

Open Research Online

The Open University's repository of research publications
and other research outputs

Molecular basis of the human ApoAII exon 3 splicing Thesis

How to cite:

Arrisi Mercado, Pablo (2004). Molecular basis of the human ApoAII exon 3 splicing. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2004 Pablo Arrisi Mercado

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000f583>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Molecular basis of the human ApoAII exon 3 splicing

Pablo Arrisi Mercado

**A Thesis Submitted in Fulfilment of the Requirements of the Open University, London
(U.K.) for the Degree of Doctor of Philosophy**

Life Sciences

International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy

Director of Studies: Francisco E. Baralle M.D., Ph.D.

Second Supervisor: Christopher W. J. Smith, Ph.D.

September 2003

-11
Submission date: 30 September 2003
Award date: 26 January 2004

ProQuest Number:27532731

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 27532731

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

A mi esposa
To my wife

Acknowledgments

Having finished this long process there are lot of people who I wish to thank, not only for their constant support in the scientific field but also as human beings. Becoming part of this group, I learned not only how exciting research is and to ask yourself to look for solutions or explanations to the problems that you have to deal with, but also how important are people that are with you and that are be able to take you away from bad moments, just with a word.

I know this is just the beginning, a necessary cycle in order to go ahead. There is a long way to go, however the bases that I have acquired here will be extremely useful for the future.

First and foremost anything I thank my wife Paola, for her sacrifice in coming to Italy leaving behind her parents, friends and our country and whose company and support have been fundamental during these four years.

I also thank the ICGEB as institution providing the great opportunity for students from the member countries to perform research and the chance to meet and interact with people from different countries and cultures.

I am deeply indebted to "Prof", Francisco Baralle for giving me the opportunity to work in his group, for the comprehensiveness during hard times and enthusiasm in good times. To my second supervisor Chris Smith from Cambridge University, for his collaboration, advice and useful discussion of this thesis.

A special thanks goes to "Zio" Mauri for his help, support and friendship during all of these years in which good and bad times were positively lived together.

I say also thanks to Roby, Emy, Andres, Mauri G, Cristiana, Marco and Franco for their willingness in the discussion, teaching, help in the technical approaches of my thesis and specially for their friendship during all of these years.

Last, but not the least, thanks to my family for their constant love and support.

TABLE OF CONTENTS

TABLE OF CONTENTS	2
LIST OF FIGURES.....	7
ABSTRACT.....	12
ABBREVIATIONS	15
1. INTRODUCTION.....	17
Pre-mRNA SPLICING.....	17
Overview of pre-mRNA processing.....	17
Chemistry of splicing.....	18
Splicing is a part of a co-transcriptional process	19
The Spliceosomal Complex	21
General model for spliceosome assembly and spliceosome cycle.....	22
CONSTITUTIVE SPLICING.....	25
Cis-acting elements.....	25
Exon/intron architectural requirements	26
The 5' and 3' splice sites requirements	26
The polypyrimidine tract	28
The branch site.....	29
Relative strength and cooperation between signals.....	29
Enhancers and silencers.....	30
The RNA secondary structure	32
TRANS-ACTING FACTORS	33
Protein components involved in splicing	33
Small Nuclear Ribonucleoprotein Particles (snRNPs)	33
hnRNP proteins.....	35
SR proteins.....	37

<i>Splicing-dependent exon-exon junction complex</i>	<i>40</i>
<i>Splice junction pairing: Exon definition and Intron definition.....</i>	<i>43</i>
<i>Alternative splicing.....</i>	<i>46</i>
The human Apolipoprotein AII exon 3 as a model to study the influence of atypical polypyrimidine tract (GU) on splicing.....	47
<i>The Apolipoprotein AII gene.....</i>	<i>47</i>
<i>Polymorphic (GT)_n within the 3' splice site of human ApoAII intron 2</i>	<i>49</i>
AIM of the RESEARCH	52
2. RESULTS	53
Analysis of the ApoAII exon 3 splice sites	53
“in vivo” and “in vitro” systems for the study of ApoAII exon 3 splicing	58
Influence of the 3' splice site strength on splicing of ApoAII exon 3.....	61
<i>In vivo effects of (GT)₁₆ tract deletion and replacement.....</i>	<i>61</i>
<i>A 45 kDa protein interacts with the 3' splice site of ApoAII exon 3.....</i>	<i>63</i>
5' splice site involvement in the recognition of the ApoAII exon 3	67
Mapping of regulatory elements within ApoAII exon 3.....	69
Screening of human ApoAII exon 3 for SR proteins consensus sequences.....	73
Interaction of SR proteins with ApoAII 9nt-ESE	76
Effect of SR proteins and hnRNP A1 overexpression on splicing of ApoAII exon 3	80
The position of ApoAII ESE within an internal exon is crucial for its function.....	83
Order of intron removal in the ApoAII gene.....	88
Mapping of regulatory elements within intronic regions flanking ApoAII exon 3	89
3. DISCUSSION	95
An Evolutionary point of view.....	95

Correlations among different sequence features of ApoAII exon 3 and CFTR exon 9	98
Influence of the 3' splice site on splicing of ApoAII exon 3	100
5' splice site involvement in the recognition of the ApoAII exon 3	102
ApoAII exon 3 sequences and splicing regulation	104
<i>Trans acting factors and ApoAII exon 3 splicing</i>	107
Secondary structure	110
The position of the ESE within ApoAII exon 3 is crucial	113
Mapping of cis-acting elements within intronic regions flanking ApoAII exon 3	114
Models of human ApoAII exon 3 constitutive splicing	117
<i>Future directions</i>	120
4. MATERIALS AND METHODS	122
Chemical reagents	122
Standard solutions	122
Enzymes	122
Synthetic oligonucleotides	123
Radioactive isotopes	123
Bacterial culture	123
Cell culture	123
DNA preparation	123
<i>Small scale preparation of plasmid DNA from bacterial cultures</i>	123
<i>Large scale preparations of plasmid DNA from bacterial cultures</i>	124
<i>Preparation of genomic DNA</i>	124
RNA preparation from cultured cells	125

Estimation of nucleic acid concentration.....	125
Enzymatic modification of DNA	126
<i>Restriction enzymes</i>	126
<i>Large fragment of E. coli Polymerase I and T4 Polynucleotide Kinase</i>	126
<i>T4 DNA ligase</i>	127
Agarose gel electrophoresis of DNA	127
Elution and purification of DNA fragments from agarose gels.....	128
Preparation of bacterial competent cells.....	128
Transformation of bacteria.....	129
Amplification of selected DNA fragments.....	129
Sequence analysis for cloning purpose.....	130
Reconstruction of the whole Apo AII gene sequence	130
Oligos sequence of the different construct	131
<i>Generation of CF/Apo hybrids</i>	131
<i>Generation of overlapping deletion within Apo AII exon 3</i>	131
<i>Generation of point mutation within Apo AII exon 3 ESE</i>	132
<i>Generation of mutagenesis of G runs in intron 3</i>	132
<i>Generation of Apo AII exon 3 5'splice site mutants</i>	133
Replacement of EDA pY tract by GT repeats	133
Intron removal order experiment	134
Oligo antisense for TDP-43 (TIO1318).....	134
Oligos for Electro Mobility Shift Assay (EMSA).....	134
Oligos for UV crosslinking	135
Oligos for UV crosslinking (ISE within intron 3).....	135

Construction of the minigene system.....	136
<i>Apo AII exon 2-intron 2-exon 3 minigene used for UV-crosslinking</i>	136
Maintenance and analysis of cells in culture	136
Transfections	137
Cotransfections	137
mRNA Analysis by Polymerase Chain Reaction	137
<i>cDNA synthesis</i>	137
<i>cDNA analysis</i>	138
Denaturing polyacrylamide gel electrophoresis (SDS-PAGE).....	139
RNA preparation	139
RNA EMSA.....	140
UV cross-linking	141
Immunoprecipitation of SR proteins following UV-crosslinking.	142
In vitro splicing assays	144
REFERENCES	145

LIST OF FIGURES

	Page
Figure 1.1. Transcriptional coupling of pre-mRNA processing	20
Figure 1.2 The spliceosome cycle is shown	23
Figure 1.3 Schematic representation of exon-intron boundaries and consensus sequences for 5' and 3' splice sites and branch point	25
Figure 1.4 Splice signal motifs in different organisms	27
Figure 1.5 Models of SR protein action in exonic-splicing-enhancer-dependent splicing	39
Figure 1.6 Exon definition versus Intron definition	44
Figure 1.7 Schematic representation of the <i>apoA-II</i> gene	48
Figure 1.8 Comparison of splice sites from the human Apo AII exon 3, human CFTR exon 9, human α cardiac actin and mouse CFTR exon 9 with consensus sequences for 5' and 3' splice sites	49
Figure 2.1 Comparison of splice sites from the mouse and human ApoAII exon 3 with human CFTR exon 9	54
Figure 2.2 Human CFTR exon 9 splice site prediction by Neural Network	55

	Page
Figure 2.3 Mouse ApoAII exon 3 splice site prediction by Neural Network	56
Figure 2.4 Human ApoAII exon 3 splice site prediction by Neural Network	57
Figure 2.5 Schematic representation of pApo construct	59
Figure 2.6 Endogenous Apolipoprotein AII and pApo-wt construct splicing pattern	60
Figure 2.7 In vivo effect of GT deletion/replacement	61
Figure 2.8 Effects of ApoAII-(GT) ₁₆ replacement within the polypyrimidine tract of mouse EDA exon	62
Figure 2.9 Effect of the GT tract deletion on RNA/protein interaction	66
Figure 2.10 Effect of ApoAII exon 3 5' splice site mutation on Apo exon 3 splicing	67

	Page
Figure 2.11 Effect of the human ApoAII 5' splice site replacement by mouse ApoAII 5' splice site and human CFTR 5' splice site	69
Figure 2.12 Effect of the progressive Apo AII exon 3 replacement by CFTR exon 9	70
Figure 2.13 Effect of the overlapping deletions within Apo AII exon 3	72
Figure 2.14 Effect of point mutations within the Apo AII exon 3 putative ESE	73
Figure 2.15 High-score SR protein motifs in exon 3 of Apo AII	75
Figure 2.16 Electro mobility shift assay of ApoAII exon 3 ESE	76
Figure 2.17 Effect of ESE deletion or mutation on RNA/protein interaction	77
Figure 2.18 Specific interaction of SR proteins with the ESE within Apo AII exon 3	79
Figure 2.19 Effect of overexpression of some SR proteins and hnRNPA1 on Apo AII exon 3 splicing	82

	Page
Figure 2.20 <i>In vitro</i> study of the effect of A97T mutation on the Apo AII exon 3 ESE	83
Figure 2.21 Relevance of the Apo AII exon 3 ESE position in <i>in vivo</i> studies	85
Figure 2.22 Apo AII exon3 ESE works only when located within internal exons	87
Figure 2.23 Splicing order of ApoAII introns 1, 2 and 3	89
Figure 2.24 Effect of G runs mutations within Apo AII exon 3 flanking intron	90
Figure 2.25 Effect of simultaneous deletions of intron 2 and 3 on Apo AII exon 3 inclusion	91
Figure 2.26 Effect of single deletion either of intron 2 or intron 3 on Apo AII exon 3 splicing	92
Figure 2.27 Effect of restoring intron 2 sequence on Apo AII exon 3 splicing	93
Figure 2.28 Effect of poly purine sequence deletion in intron 3 on Apo AII exon 3 splicing	94

	Page
Figure 3.1 Comparison of ApoAII genes identity among different species	95
Figure 3.2 Comparison of ApoAII exon 3 in different species (Rat, Human, Mouse, Macaca fascicularia and Chimpanzee)	97
Figure 3.3 Lowest free-energy structure of the ApoAII exon 3 wild-type (WT)	111
Figure 3.4 Scheme of the cis-acting elements and trans acting factors involved in the human ApoAII exon 3 definition	119
Figure 4.1 Electrophoretic Mobility Shift Assay (EMSA)	140
Figure 4.2 Scheme of UV cross-linking assay	141
Figure 4.3 UV cross-linking/Immunoprecipitation assay	143

ABSTRACT

Among human genes, CFTR intron VIII/exon 9 and apolipoprotein AII intron II/exon 3 boundaries share the characteristic feature given by the presence of a peculiar tract of alternating pyrimidines and purines close to the 3' splice site.

In the case of CFTR gene, the pyrimidine rich tract is composed by a stretch of Ts in a row and a stretch of alternating pyrimidines and purines (microsatellite TG dinucleotide repeats) and both tracts are polymorphic for their length. On the other hand, in the case of ApoAII gene the pyrimidine rich tract is made exclusively of alternating pyrimidines and purines (microsatellite TG dinucleotide repeats) and it is also polymorphic for its length in both genes. This apparent sequence similarity, concerning the TG tract, is contrasted by the different splicing pattern exhibited by the two genes. In fact, CFTR exon 9 undergoes alternative splicing to a variable extent, depending on the variations in length of the pyrimidine rich tract, whereas ApoAII exon 3 is constitutively included in mRNA. Previous studies of our group have shown in the CFTR intron VIII/exon 9 context, the stretch of pyrimidines alternated with purines (within the UG tract) alone is not equivalent to a functional continuous polypyrimidine tract, contrarily to what has been observed for the apolipoprotein AII gene. Moreover, the comparison of splice sites strength of human ApoAII exon 3 and human CFTR exon 9 has outlined an apparent contradiction between the splicing behavior of the two exons and the strength of the splice sites. In fact, both the 3' and the 5' splice sites of CFTR exon 9 display a good match with the consensus whereas the match of the ApoAII exon 3 splice sites is not good. Altogether these observations prompted us to investigate the mechanisms underlying the constitutive splicing of ApoAII exon 3 and, in particular, to characterize the cis-acting elements and the trans-acting factors involved in ApoAII exon 3 definition to assure its constitutive splicing.

In order to study *in vivo* the splicing mechanism of ApoAII exon 3, we set up an eukaryotic expression system by cloning the whole ApoAII gene, from its promoter to the poly-A signal. Then, the effects of point mutations, deletions or substitutions on splicing of exon 3 were analyzed by RT-PCR after transient transfection in Hep3B cell line.

Deletion or replacement of the UG repeats at the 3' splice site of intron 2 resulted in a significant increase in exon 3 skipping, indicating the importance of this alternated arrangement of U and G as a functional polypyrimidine tract or at least as an important sequence able to lead the exon 3 definition. Furthermore, UV-crosslinking assays showed that the (UG)₁₆ repeats of ApoAII intron 2 are recognized by TDP-43, a protein that binds specifically the UG tract within the context of the 3' end of CFTR intron VIII and that affects negatively CFTR exon 9 splicing. Next, we characterized the exonic cis-acting elements able to affect the splicing efficiency of ApoAII exon 3. Transient transfections of different constructs of the ApoAII gene system carrying deletions or point mutations showed that the region spanning from nucleotide 87 to 113 of human ApoAII exon 3 is important for its inclusion in mRNA. In particular, we mapped a core enhancer-like 9nt-sequence (ESEwt) close to the 5' splice site. In order to identify trans-acting factor/s able to bind the 9nt-ESE within ApoAII exon 3, both Electro Mobility Shift Assay (EMSA) and UV-crosslinking coupled to immunoprecipitation assays were carried out. EMSA showed a broad band of shifted material with the ESEwt RNA, whereas no significant shift was seen with mutated ESE RNA. This suggested that one or more proteins interact specifically with the wild type ApoAII exon 3 across the 9nt-sequence. Then, UV-crosslinking followed by immunoprecipitation with monoclonal antibodies anti-SR proteins ASF/SF2 and anti-SC-35 showed that ESEwt but not mutated ESE RNA was able to immunoprecipitate a band whose molecular weight corresponds to that of ASF/SF2 and, even if to a lower extent, also a band

whose molecular weight corresponds to that of SC35. Thus, these results provided an evidence that at least two SR proteins are able to interact with the sequence across the ESE sequence. Subsequently, the relevance of ESE position within an internal intron (and therefore with other flanking cis-acting elements) was also tested. Both *in vitro* and *in vivo* experiments showed that the ApoAII exon 3 ESE works only if it is localized within an internal exon and not if it is placed within the first or the last exon. These results also suggested the presence of other splicing regulatory elements within the flanking intronic regions of ApoAII exon 3. Thus, to explore intron 2 and 3 for the presence of cis-acting elements able to affect the splicing of exon 3, a series of deletions within both introns were carried out. Thus, we found at least one regulatory element placed within intron 3 that regulates positively exon 3 inclusion. In conclusion, the constitutive splicing of ApoAII exon 3 seems to be the result of the balance between positive and negative action of the regulatory elements found in the exon 3 and its flanking introns. Future studies will be aimed at identifying the factors interacting with the intronic regulatory elements and at defining a possible model to explain the mechanism of ApoAII exon 3 constitutive splicing by integrating the network of interactions among the identified cis-acting elements and trans-acting factors.

ABBREVIATIONS

The standard abbreviations used in this dissertation follow IUPAC rules. All the abbreviations are defined also in the text when they are introduced for the first time. The abbreviations mentioned only once are not included in this list.

bp	Base pairs
nt	Nucleotides
aa	Amino acid
kb	Kilobase
kDa	Kilodalton
dNTPs	Deoxynucleoside triphosphate (A, C, G and T)
N	Nucleotide (A or C or G or T)
DNA	Deoxyribonucleic acid
cDNA	Copy DNA
ESE	Exonic Splicing Enhancer
ISE	Intronic Splicing Enhancer
ESS	Exonic Splicing Silencer
ISS	Intronic Splicing Silencer
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein particles
hnRNP	Heterogenous ribonuclear protein
SR	Arginine-serine rich protein
ss	Splice site
NMD	Nonsense-mediated decay

DTT	Dithiothreitol
EDTA	Ethylenediamine tetra-acetic acid
IPTG	Isopropyl-b-d-thiogalactopyranoside
TBE	Tris-borate-EDTA (buffer)
PBS	Phosphate buffer saline
ddH ₂ O	Double-distilled water
Pu	Purine

1. INTRODUCTION

Pre-mRNA SPLICING

Overview of pre-mRNA processing

Eukaryotes genes are often interrupted by intervening sequences (introns or IVS) that must be removed during gene expression. RNA splicing is the process by which these intervening sequences are precisely removed and the flanking, functional sequences (exons) are joined together. Therefore, RNA splicing is one of the major steps in the control of gene expression in eukaryotes.

Pre mRNA splicing is one of the latest stages recognized in the flow of genetic information from gene to protein. The presence of introns was first described in the non-coding regions of adenovirus (Berget et al., 1977; Chow et al., 1977) but was quickly shown to be a common feature of cellular genes. In fact, in 1977 Jeffreys and Flavell described the presence of “large insert” in the coding sequence of rabbit beta-globin gene. In 1978, Chambon and his colleagues reported the positions where the coding sequences for ovalbumin mRNA were interrupted in the genome (Breathnach et al., 1978). Moreover, they found repeated sequences at exon-intron boundaries that share common features, and this led to the proposal that there are unique excision-ligation points common to all boundaries (Breathnach et al., 1978).

In pre-mRNA splicing, introns are removed and exons are ligated together by a two-step transesterification reactions carried out by the spliceosome, a dynamic 60S ribonucleoprotein particle (Staley and Guthrie, 1998). Formation of the spliceosome at particular splice junctions is triggered by recognition of the 5' splice site by the U1 snRNP and of the 3' splice site by U2AF, followed by the U2 snRNP. It is unclear whether the spliceosome is assembled from larger complexes, such as the recently identified penta-snRNP, which contains the U1, U2,

U4, U5 and U6 snRNAs (Stevens et al., 2002) or the 200S InRNP (large nuclear ribonucleoprotein particle), which contains additional non-snRNP RNA processing factors (Yitzhaki et al., 1996), or by the sequential addition of snRNP and non-snRNP factors as was previously supposed

Chemistry of splicing

From a chemical point of view, the reactions involved in pre-mRNA splicing are quite simple: the exchange of one substituent on a phosphodiester bond for another. Although the intermediates and products of splicing were characterized some years ago, the reaction mechanisms have been established with certainty only recently, and the pre-mRNA structural features that are important for the transesterification reactions are just now being investigated. If the reactions are protein- or RNA- catalyzed also remains an open question.

Splicing occurs in two steps involving two sequential transesterification reactions (Moore et al., 1993). The first step involves cleavage at the 5' splice site to yield a 5' exon intermediate (exon1) with a free 3'-OH terminus. Concurrently, the 5' end of the intron (IVS) is joined, via a 2'-5' phosphodiester bond, to an adenosine residue within the intron (the branch site) generally located within 100 nucleotides of the 3' splice site. This forms the so-called lariat intermediate RNA containing a branched intron with attached 3' exon (exon 2), The exon 1 and IVS-exon 2 intermediates are subsequently resolved by cleavage at the 3 splice site and ligation of the two exons via a 3'-5' phosphodiester bond. This yields two products: spliced exons and excised lariat intron with a 3'-OH terminus.

Although ATP hydrolysis is required for intron excision from pre-mRNA, none of the phosphates in the spliced products is derived from exogenous ATP. Rather the splice site phosphates are conserved in the products (Konarska et al., 1985; Lin et al., 1985; Padgett et al., 1984). In addition, it had been found that neither step of pre mRNA splicing could be

separated into bond- breaking and bond- forming reactions. Consistent with these observations, it has been shown that both steps proceed via direct “in line” SN2 displacement mechanism. Thus the first step involves direct nucleophilic attack by the 2' OH of the branch site adenosine on the phosphate at the 5' splice site. This displaces the 3' oxygen of the 5' exon by way of a pentacoordinate transition state. In the second step, the same 3' OH of the 5' exon, which had been the leaving group in the first reaction, reverses its role and acts as a nucleophile at the 3' splice site phosphate to similarly displace the 3' end of the intron. Interestingly, these direct displacement reactions parallels the known reaction mechanism of group I self-splicing introns and are therefore consistent with the notion that pre-mRNA splicing is an RNA-catalyzed process.

Splicing is a part of a co-transcriptional process

From the point of view of RNA Polymerase II (Pol II), the transcription process includes pre-initiation complex formation, transcription initiation, elongation, termination and dissociation of Pol II from the DNA template (Fig.1.1). The transcript also undergoes pre-mRNA processing steps including: (1) 5' end capping, in which the 5' triphosphate of the pre-mRNA is cleaved and a guanosine monophosphate is added and subsequently methylated to produce m7GpppN; (2) editing, in which individual pre-mRNA residues are converted to alternative bases (e.g. adenosine is converted to inosine by base deamination) to produce mRNAs encoding distinct protein products; (3) splicing, in which introns are removed and exons are ligated together by the spliceosome; (4) 3' end formation, which involves pre-mRNA cleavage and synthesis of the poly(A) tail; and paradoxically (5) degradation.

A priori, each of these modifications might occur independently of the others, since most can occur in *in vitro* reconstituted systems on purified pre-mRNA substrates. However, many studies have revealed functional relationships among these processes and each (with the

exception of editing) has been shown to be co-transcriptional, at least some of the time. Importantly, a number of trans-acting factors required for pre-mRNA processing directly bind to Pol II, which stimulates processing, and, in some cases, processing feeds back to Pol II activity. This has led to the proposal that RNA transcription and processing occur in a “gene expression factory” composed of machines linked together for the purposes of efficiency and regulation (Bentley, 2002; Cook, 1999; Maniatis and Reed, 2002; Proudfoot et al., 2002).

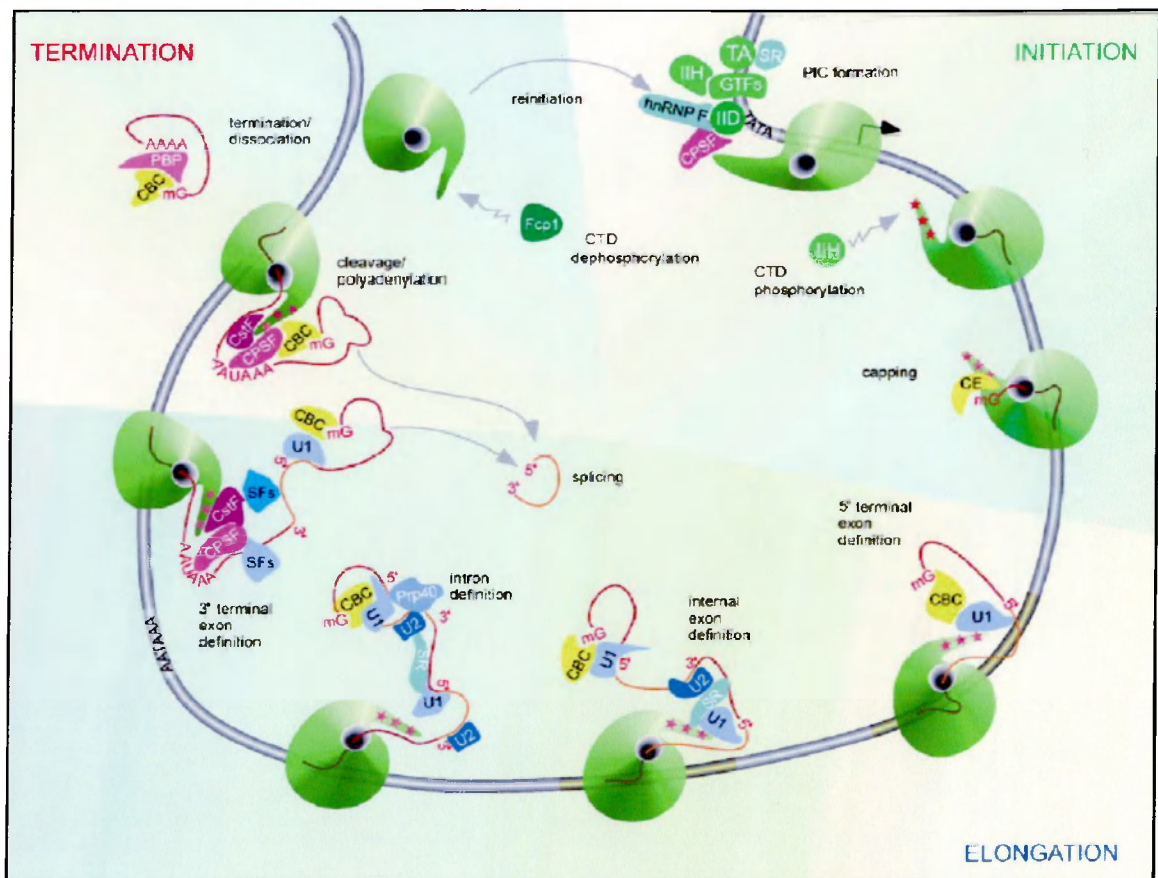


Figure 1.1. Transcriptional coupling of pre-mRNA processing. The transcription cycle is depicted starting at the top right and proceeding in a clock-wise fashion. Processing factors may be recruited to the PIC at the promoter (TATA) by GTFs (TFIID, IID; TFIIF, IIF) and transcriptional activators (TAs). Hypophosphorylated RNAP II (green “tadpole” depicted with a CTD “tail”). IIF phosphorylates the CTD (red stars) soon after initiation, which recruits CE. Associations retained between the CTD and various factors during elongation are not shown for simplicity. CBC mediates 5Vterminal exon definition. IVSs are denoted by the gold DNA. SR proteins and the CTD facilitate internal exon definition. Intron definition occurs when splice sites are paired across introns (Prp40 is shown assisting this process). 3V-End processing is mediated by cleavage/polyadenylation factors assembled at the poly(A) site (AATAAA) and is assisted by the CBC and splicing factors (SFs). These interactions mediate 3Vterminal exon definition. Transcription termination is accompanied by CTD dephosphorylation (proposed to occur by Fcp1). Figure taken from Howe et al (Howe, 2002)

By the time, pre-mRNAs (hnRNAs) emerge from the transcription complex, and throughout the time they are in the nucleus they are associated with proteins. The most abundant cellular proteins that bind hnRNA are the heterogeneous nuclear ribonucleoprotein (hnRNP) that together, formed the hnRNP particles [reviewed in (Dreyfuss et al., 1993)]. Some members of this family of proteins have helix-unwinding activity and seem to function primarily to maintain the transcript in a single-stranded state. They influence the structure of hnRNA thus facilitating the interaction of hnRNA sequences with the other components that are needed for its processing event. hnRNP proteins may also play important roles in the interaction of hnRNA with other nuclear structures, in nucleocytoplasmic transport of mRNA, and in other cellular processes (Visa et al., 1996). The dynamic structures within the nuclei, called speckles, are places of localization of many transcription and splicing factors. (Zeng et al., 1997). Upon the activation of the genes, those factors are recruited to the sites of transcription, bind to the hnRNA in the highly ordered cooperative way, thus driving a processing event and then return to the speckles between rounds of active pre-mRNA synthesis for recycling and reactivation (Sleeman and Lamond, 1999). The cooperativity is shown for transcription and splicing processes which occur contemporarily and influence each other (reviewed in (Bentley, 1999; Corden and Patturajan, 1997)).

The Spliceosomal Complex

Two of the main functions of the spliceosomal snRNPs are to recognize the 5' and 3' intron/exon boundaries and to assemble onto these sites the macromolecular enzyme that catalyze the splicing reaction. In mammals four distinct spliceosomal complexes, which form in the temporal order E, A, B and C, have been detected (Reed, 1997) (Figure 1.2). These general steps in spliceosomal assembly are now well understood, with some variations in the model proposed few years ago.

General model for spliceosome assembly and spliceosome cycle

Assembly of the major spliceosome is initiated by the ATP-independent recognition of the 5' splice site by the U1 snRNP, which leads to the formation of the spliceosomal complex E. This interaction is mediated by base pairing of the U1 snRNA with the 5' splice site, as well as by protein-protein and protein-pre-mRNA interaction involving U1-70K and U1-C proteins (Will, 1997). Novel insights into the spliceosomal assembly are coming from studies in *S. cerevisiae* in which has been shown that the first proteins that interact with the 5' splice site during the commitment complex (E complex counterpart) formation are the cap-binding complex protein CBC80 and seven U1 snRNP (Zhang and Rosbash, 1999). Metazoan homologues of some of these proteins have been identified, indicating that similar interaction may occur in E complex. One of these proteins is Nam8 or TIA1 in metazoan, which affects splice-site recognition (Puig et al., 1999).

Another important event that occurs during formation of the E complex is the recognition of the 3' end of the intron. In higher eukaryotes, the 3' splice site and adjacent pyrimidine tract are identified through interactions with a dimeric splicing factor U2AF. This, in turn, helps recruitment of the U2 snRNP to the branchpoint together with a branchpoint binding protein (BBP or SF1).

The current model of assembly proposes that U2 snRNP first associate in an ATP-dependent manner in the A complex. However, U2 snRNP has now been identified as a component of a purified, functional E complex. U2 snRNP association with the E complex occurs in the absence of ATP and does not require branch point sequence interaction. The interpretation of this data is that U2 snRNP first binds loosely to the pre-mRNA in the E complex via the integral U2-snRNP-associated protein SF3b, and then an ATP- dependent process leads to stable binding of U2-snRNP to the branch point sequence in the A complex and this is facilitated by the presence of other proteins such as SF3A and SF1 (Das et al., 2000).

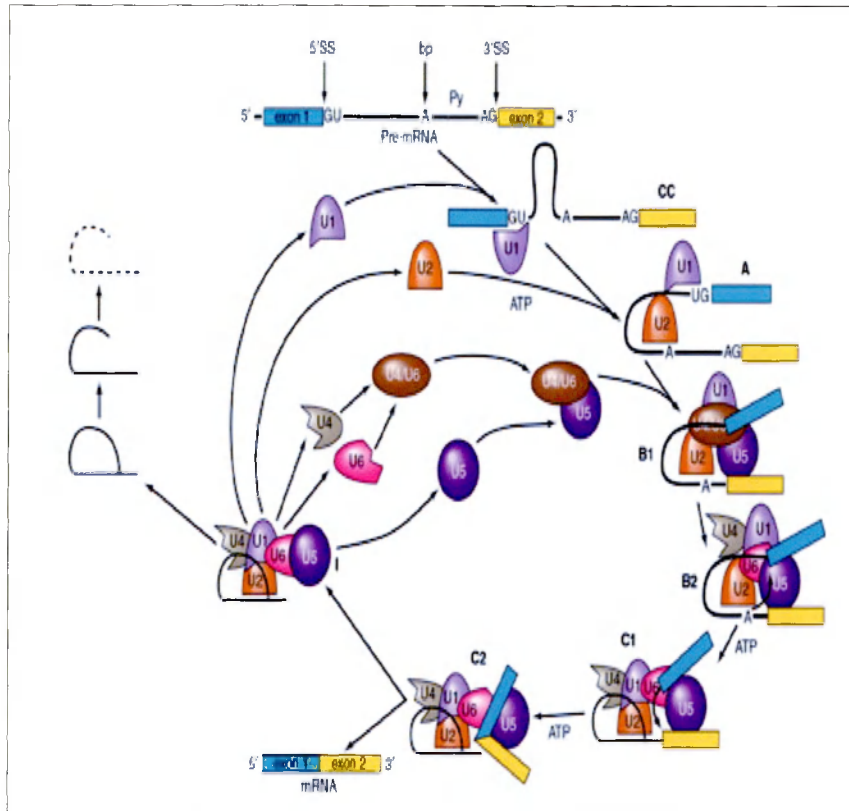


Figure 1.2 The spliceosome cycle is shown. The figure summarizes each of the steps, in the order that they occur, and the energy requirement, if any. There are 5 snRNPs involved: U1, U2, U4, U5 and U6 (adapted from Moore et al. 1993).

The next major event after A complex assembly is the binding of U4/U5•U6 tri-snRNPs at the 5' splice site to form the B complex. At this time many structural snRNPs rearrangements occur leading to the formation of a complex RNA-RNA network within the spliceosome. U4/U6 base pairing interaction is disrupted and U6 snRNA base pairs with U2 snRNA and also the 5' splice site. Furthermore, U5 snRNA loop I base pairs with exon sequence at the 5' splice site and later with exon sequences at 3' splice site. The 5' and 3' splice sites are also contacted by the U5-220 KDa protein (Umen and Guthrie, 1995), which has been reported to affect also the tertiary interaction between the splice sites and U6 snRNA (Collins and Guthrie, 1999). Recent studies report that tri-snRNPs interact with the 5' splice site and the upstream 5' exon at earlier step in spliceosome assembly than previously thought (Maroney et al., 2000). These interactions occur in an ATP-dependent manner in the absence of stable U2

snRNP binding and appear to be guided, in part, by the U5 associated protein U5-220KDa. In this view, formation of B complex does not reflect “recruitment” but rather a stabilization that occurs as a consequence of interspliceosomal RNA-RNA rearrangements such as replacements of U1 by U6 and formation of U2/U6 helices that are prerequisite for the formation of the catalytically competent complex C.

An unsolved problem is the catalysis of RNA in pre-mRNA splicing, although is widely believed that it is mediated by RNA. Spliceosome is a metallo-enzyme (Sontheimer et al., 1997; Steitz and Steitz, 1993) and from recent studies has been shown that U6 snRNA coordinates metal ion thus contributing to pre-mRNA splicing (Yean et al., 2000). U6 is a good candidate for catalysis because is highly conserved through evolution and also forms intramolecular and intermolecular helices that are analogous to auto catalytic group II introns (Yu et al., 1995).

CONSTITUTIVE SPLICING

Cis-acting elements

The question of splice-site choice is closely connected to the problem of normal recognition of constitutive splice sites. A feature shared by both regulatory sequences and splice-site signals is that they are usually short and often degenerate. Cis-acting determinants that influence competing splicing pathways include relative strength of 5' splice site, the branch point sequence, the polypyrimidine tract, the proximity between 5' splice site and branch point, sequences between the branch point-polypyrimidine tract and 3' splice site (Figure 1.3). In some cases, the formation of pre-mRNA secondary structure participates in the regulation of splice site selection by modifying the physical distance within introns, or being involved in the definition of the exon. Moreover, additional exonic or intronic sequences are often involved in the correct recognition of splice sites, when sub optimal splice sites are present. These sequences can act by increasing the recognition or decreasing it and are respectively named enhancers or silencers.

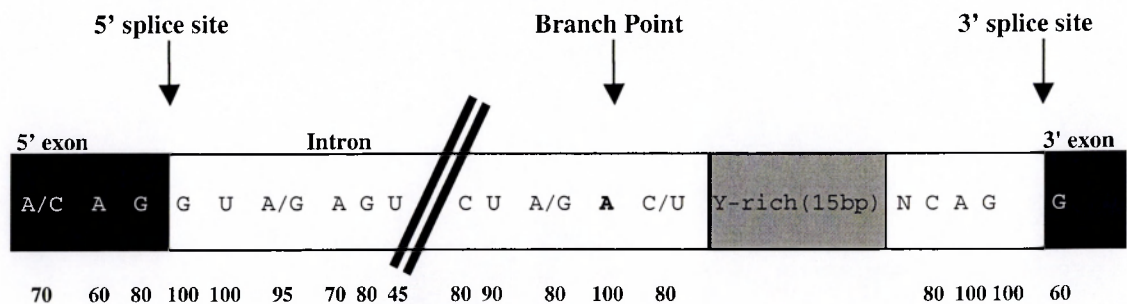


Figure 1.3 Schematic representation of exon-intron boundaries and consensus sequences for 5' and 3' splice sites and branch point. Black boxes are exons, Y indicates a pyrimidine and N means any base. Yn means region of about 15 pyrimidines. The bolded A is the branch point. The frequencies of each nucleotide in a specific position are indicated.

Exon/intron architectural requirements

In most cases, vertebrate genes consist of multiple small exons separated by intervening sequences that can be significantly larger. The average size of exons is 137 nucleotides and exons longer than 300 nucleotides or shorter than 50 nucleotides, even existing, seems to be not favored by splicing machinery (Hawkins, 1988). Artificially shortening an internal exon leads to inefficient recognition (Dominski and Kole, 1991; Dominski and Kole, 1992) presumably due to both deletion of exon accessory sequences and steric hindrance of factors such as U1 and U2 snRNPs.

Moreover, a compensatory relationship between exon and intron size has been proposed. In fact, Sterner *et al.* observed that large internal exons are problematic for recognition if they are flanked by large introns, suggested that naturally occurring large vertebrates exons might be flanked by small introns (Sterner et al., 1996).

The 5' and 3' splice sites requirements

Splice site strength is an important determinant in splice site selection (fig.1.4). Consensus sequences are limited to 4 bases at the 3' intron/exon junction and to 9 bases at the 5' exon/intron junction (Shapiro and Senapathy, 1987). It is shown that mutating one of the first two bases (AG) in intron, immediately upstream of 3' splice site (YAG/G), as well as GU in intron, immediately downstream of 5' splice site (A/C)AG/GUPuAGU) completely abolish splicing (Langford et al., 1984). Studies of alternative splicing have shown that mutations that improve the match of weak splice sites to the consensus can led to the constitutive recognition of alternatively skipped exon (Del Gatto and Breathnach, 1995; Huh and Hynes, 1993; Muro et al., 1998), also the use of the normal 3' splice site can be restored either improving the match of a cryptic branch site to the branch site consensus or by introducing mutant U2 snRNAs with greater complementarity to the cryptic branch site (Zhuang and Weiner, 1989).

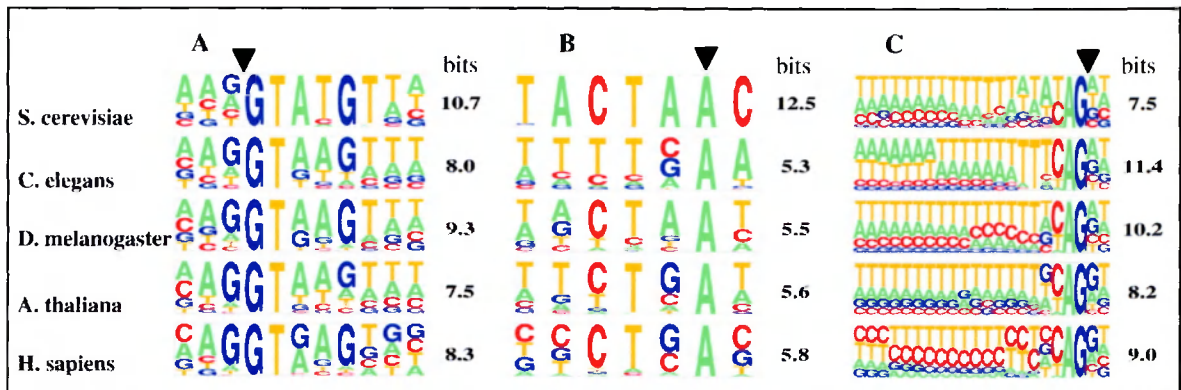


Figure 1.4 Splice signal motifs in different organisms. Sequence motifs for the 5'ss (A), branch site (B), and 3'ss (C) are displayed. The height of each letter is proportional to the frequency of the corresponding base at the given position, and bases are listed in descending order of frequency from top to bottom. The RelEnt (in bits) of the motif model used relative to the background transcript base composition is also shown. The splice junctions and branch point are marked by inverted triangles. Figure taken from Lim et al (Lim and Burge, 2001)

Mutations affecting splice sites are common in human genetic disease. In fact, a survey performed more than a decade ago found that 15% of point mutants that result in human genetic disease affect pre-mRNA splicing (Krawczak et al., 1992). The majority of these are single-point mutations affecting the conserved bases at the donor or acceptor splice sites. (Cooper DN, 1995).

A study of more than a hundred splice-site mutations showed that point mutations affecting the 5' donor splice site were more common than those at the 3' acceptor site (62% vs 26%). At the 5' donor splice site, mutations affecting the G residue at position +1 are the most common, followed by mutations at position +5. Mutations at these two positions are thought to significantly reduce the pairing of the donor splice site with the complementary site in the small nuclear ribonucleoprotein particle U1snRNP, which is one of the first steps in the complex process of pre mRNA splicing (Kramer, 1996; Krawczak et al., 1992; Wittop Koning and Schumperli, 1994)

At the 3' acceptor site, mutations affecting the conserved -2 and -1 sites (A and G dinucleotides respectively) are most common, although mutations affecting the +1 site are also

observed (Ariffin et al., 2003; Ikeda et al., 2001). Mutations at donor or acceptor sites can lead to exon skipping, intron retention or insertions and deletions due to utilization of cryptic splice sites.

The polypyrimidine tract

The adjacent polypyrimidine tract can affect the recognition of the branch point sequences in the metazoan (Mullen et al., 1991; Reed, 1989). It has been shown that progressive deletion of the polypyrimidine tract abolish lariat formation, spliceosome assembly and splicing (Mullen et al., 1991; Roscigno et al., 1993). The polypyrimidine tract is one of the important cis-acting sequence elements directing intron removal in pre-mRNA splicing. It is known that polypyrimidine tract deletion or mutation decreases the splicing efficiency as well as the elongated length of pyrimidine in row can led to improved efficiency of splicing (Roscigno et al., 1993).

Despite the important role of polypyrimidine tract in splicing, there appears to be a great flexibility in the specific sequence of a given tract.

Further studies have demonstrated by cis-competition assay that the polypyrimidine strength is not only determined by its length but also by its composition (Coolidge et al., 1997). The best polypyrimidine tract is composed by a poly-uridine tract, whereas, a poly-cytidine tract is not functional (Roscigno et al., 1993). Then, proximity of the polypyrimidine tract to the 3' splice site becomes important when the pyrimidine length is limiting. In fact, limiting number of continuous uridines to six demands that these uridines are placed immediately adjacent to the 3' splice site AG. On the other hand, a polypyrimidine tract with 11 uridine is a highly competitive pyrimidine stretch regardless of distance between the branch point and polypyrimidine tract itself (Coolidge et al., 1997). It is shown that functional pyrimidine tracts do not absolutely require continuous uridines. In constitutively included exon 3 of the human

Apolipoprotein-AII (Apo AII) the “polypyrimidine tract” is composed of a (GT)₁₆ dinucleotide repeats (Shelley et al., 1985). In this case it is likely that total uridine content may relate to the splicing efficiency.

The branch site

While in yeast there is an invariant branch point sequence (UACUAAC) which base paired with U2 snRNA, in metazoan this sequence is not highly conserved (Fig. 1.4). However, several lines of evidence have led to the belief that the mammalian branch point is specified primarily by its proximity to the 3' splice site. Most branch points have been mapped within 18-40 nucleotides of the 3' splice site (Reed and Maniatis, 1985; Ruskin et al., 1985). However there are cases in which the branch point located more than 40 nucleotides from the 3' splice site is functional and essential for the regulation of alternative splicing. In the rat α -tropomyosin gene intron 2, the branch point lies 172 nucleotides upstream from the 3' splice site. This proximity to the 5' splice site of exon 2 enforces mutually exclusive behavior of exons 2 and 3, because the splicing factors are unable to bind productively to the two elements simultaneously and form active spliceosome (Smith et al., 1989).

Generally, the mutation of the adenine residue involved in the lariat formation strongly reduces splicing efficiency of the downstream exon (Reed and Maniatis, 1988).

Relative strength and cooperation between signals

Several studies have disclosed the synergy existing either between the 5' and 3' splice sites or between the polypyrimidine tract and the branch point.

It has been demonstrated that strong sequence within the 5' splice site of an exon can promote the use of its own 3' splice site (Nasim et al., 1990). On the other hand, sequences, upstream

the 3' splice site of an exon can facilitate the use of a downstream 5' splice site (Tsukahara et al., 1994).

The polypyrimidine tract determines the location of the branch point sequence and indirectly the 3' splice site (that is usually the first AG downstream the branch point). It is also shown that a strong polypyrimidine tract can partially balance a weak branch point sequence. Likewise, a strong branch point site can partially balance a weak polypyrimidine tract (Buvoli et al., 1997). Finally, also the distance between the branch point sequence and the pyrimidine tract is critical for efficient lariat formation (Gattoni et al., 1988).

Enhancers and silencers

As previously highlighted, higher eukaryotes have large introns defined by more degenerate splicing signals with respect to yeast genes. Additional intronic and exonic sequences are often necessary for efficient and accurate choice of the correct splice site. These auxiliary elements are often conserved between species and perhaps between similarly regulated genes, but they contain degenerate sequence motifs, making it difficult to identify them. They can be exonic or intronic. When localized in introns they can lie upstream, downstream, or flanking both sides of the regulated exon. Intronic elements can also be proximal (within 100 nucleotides) or distal (more than one kilobase away from the regulated exon), although they are often located close to the exon. And finally, auxiliary elements can enhance or repress splice-site selection. Depending on their location and their effect on the recognition of alternative splice sites, the elements are referred to as exonic splicing enhancers (ESE) or silencers (ESS) or intronic splicing enhancers (ISE) or silencers (ISS).

Most splicing enhancers are located within 100 nucleotides of the 3' splice site but are not active when are located further away (Tian and Maniatis, 1994). However in *Drosophila melanogaster* doublesex pre-mRNA, the *dsx* enhancer element functions when located 300

nucleotides downstream of the regulated 3' splice site (Tian and Maniatis, 1993). Studies have demonstrated that the activity of a splicing enhancer decreases as a function of distance from the 3' splice site. Indeed, when a strong enhancer and a weak enhancer are compared, the first one has a higher ability to function at a great distance from the 3' splice site (Graveley et al., 1998).

Most enhancer elements obtain contained extended purine-rich sequences (more than 65% purine rich), but a second, novel class of sequences lacking stretches of purines was also identified (Tian and Kole, 1995). In fact, studies conducted by Schaal and Maniatis identified pyrimidine-rich enhancers that are more than 67% pyrimidine-rich and function such as strong enhancers (Schaal and Maniatis, 1999b).

In addition to sequences that promote exon inclusion, there are sequences that inhibit splicing, so called exonic or intronic splicing silencers. The silencers are less well characterized; they can be purine or pyrimidine-rich and bind a diverse array of proteins (Fairbrother and Chasin, 2000). A negative element (CAAGG) involved in the down regulation of fibronectin EDA exon inclusion was identified within EDA exon (Caputi et al., 1994), where it seems to be determinant for RNA conformation (Muro et al., 1999). An example of intronic splicing silencer is described for CFTR exon 9 alternative splicing. This extended region in intron 9 acts like silencer by recruiting SR proteins (Pagani et al., 2000).

Negative effect of silencers can be due to the interaction with protein factors. A well-known factor that inhibits splicing is PTB (polypyrimidine tract binding protein). It recognizes pyrimidine-rich elements both in introns and in exons and can function either by antagonizing U2AF65 action or creating a region of silencing across the down regulated exon (Wagner et al., 1999).

The negative effect on splicing of an inhibitory protein that is bound to a juxtaposed exonic splicing silencer (ESS) of a silencer might be antagonized by SR protein bound to an ESE (Fig. 1.3b).

The RNA secondary structure

The cis-acting elements regulating constitutive and alternative splicing interacts with trans-acting factors. The secondary structures that pre-mRNA creates might influence directly this kind of interaction.

There are many studies that proposed the RNA secondary structure as a regulator of alternative splicing. In the case of rat calcitonin/CGRP pre-mRNA splicing, the 3' splice acceptor of exon 4 forms a stable stem-loop structure. Mutations that destabilize the stem abolish the usage of this splice acceptor site in vitro (Coleman and Roesser, 1998). Also other cis-acting elements are subjected to this kind of regulation. In fact, in fibronectin EDA exon there are an ESS and an ESE. Functional studies coupled to secondary structure analysis suggest that the role of the ESS element may be exclusively to ensure the proper RNA conformation and raise the possibility that the display of the ESE in a loop position may represent a significant feature of the exon splicing regulatory region (Muro et al., 1999). Moreover, an example of 5' splice site mutation that affects a stem-loop structure can be found in the tau gene. In this case the disruption of secondary structure increase exon 10 inclusion and this result in the production of an aberrant protein isoform that leads to neuro-degeneration (Grover et al., 1999).

TRANS-ACTING FACTORS

Protein components involved in splicing

Protein components of the splicing machinery, involved in splicing event are the small nuclear ribonucleoprotein particles (snRNPs) and other non-snRNP essential splicing factors. Most of these proteins share very similar structural characteristics such as different number of RNA binding motifs and/or protein binding domains.

Small Nuclear Ribonucleoprotein Particles (snRNPs)

The snRNP particles form part of the catalytic macromolecular complex called spliceosome. Each snRNP particle consists of snRNA molecule complexed with a set of eight Sm or Sm-like proteins and several particle specific proteins (Will, 1997). The major spliceosomal snRNPs U1, U2, U4, U5 and U6 are responsible for splicing the vast majority of pre-mRNA introns (so-called U2-type introns). A group of less abundant snRNPs, U11, U12, U4atac and U6atac, together with U5, are subunits of the so-called minor spliceosome that splices a rare class of pre-mRNA introns, denoted U12-type (Burge and Sharp, 1999).

snRNAs U1, U2, U4, U5 and U6 are characterized by their small size, metabolic stability and a high degree of sequence conservation (Kambach et al., 1999). Not all the snRNA show extensive complementarity to consensus splice sites sequences, only U1 and U2, and U6 to a minor extent.

The snRNAs are transcribed by RNA polymerase II (with the only exception of U6, and presumably U6atac snRNA, that is transcribed by RNA polymerase III and entirely assembled in the nucleus). The other pre-snRNAs must be transported to the cytoplasm where snRNPs assembly with the Sm proteins is initiated and then reimported into the nucleus thanks to the

bipartite “snRNP nuclear localization signal” (NLS) formed by m3G cap and the Sm core domain (Fischer et al., 1993).

The U4 and U6 snRNAs are extensively base paired (amounting to >20 base pairs) in the U4/U6 snRNP. They associate in the nucleus forming a larger ribonucleoprotein complex. The U5 snRNA then assemble in an ATP dependent reaction with the U4/U6 snRNP giving the U4/U6•U5 three-snRNPs particle (Konarska and Sharp, 1987).

The structural core of snRNPs is formed by eight proteins, called Sm proteins, B', B, D1, D2, D3, E, F and G. This class of common proteins plays an essential role in the biogenesis of the snRNPs. The Sm proteins form three distinct heteromeric complexes prior to their interaction with the highly conserved Sm site (PuAU4-6Gpu flanked by two stem-loop structures) of the U1, U2, U4 and U5 snRNAs (Raker et al., 1996).

Sm-like proteins belonging to the Sm protein family are specifically required for the assembly of U6 snRNA. This subclass of Sm-like proteins shares the conserved structural motif characteristic of all Sm-proteins family members (Mayes et al., 1999), but can be isolated as a heteromeric complex in the absence of U6 snRNA (Achsel et al., 1999).

Besides Sm proteins, there are other particle-specific proteins that associate with snRNAs (Will, 1997). U1-70K and U1-A proteins bind directly to the RNA and are involved in the splice site recognition and selection, while U1-C associates via protein-protein interactions with U1-70K and other Sm proteins. A subset of U2 snRNP proteins also plays a critical role in tethering the U2 snRNP to the pre-mRNA. These proteins include the heteromeric splicing factors SF3a and SF3b (Brosi et al., 1993) and bind the 20-nucleotide region just upstream of the branch site in a sequence-independent manner (Champion-Arnaud and Reed, 1994; Gozani et al., 1996).

At least five different proteins associate with U4/U6 snRNP, including a 15.5 KDa protein, polypeptides of 20, 60 and 90 KDa that form complex with one another (Teigelkamp et al.,

1998). U5 snRNP particle presents a complex protein composition. U5 220KDa protein cross-links to both 5' and 3' splice sites, as well as to the exon flanking these two splice sites (Umen and Guthrie, 1995).

hnRNP proteins

Immediately after transcription, primary transcripts are associated with a set of abundant factors, called heterogeneous ribonuclear proteins (hnRNPs). hnRNP proteins are located throughout the nucleus and some of them are extremely abundant (~100 million copies per nucleus), while others are present only in small amounts (Hanamura et al., 1998; Kamma et al., 1995; Markovtsov et al., 2000; Teigelkamp et al., 1998). hnRNPs remain associated with pre-mRNA until processing is completed and with mature mRNAs during nucleocytoplasmic transport (Izaurralde and Mattaj, 1995; Michael et al., 1995). By associating with pre-mRNAs, hnRNP proteins are able to affect the interactions of pre-mRNAs with components of the RNA processing machinery (Dreyfuss et al., 1993). A subset of hnRNPs shuttles continuously between the nucleus and the cytoplasm. hnRNPs share structural properties with protein components of stable spliceosomal particles (e.g. snRNPs) or splicing regulators, since they contain common RNA-binding and auxiliary domains. Contrarily to snRNPs, hnRNPs complexes are unusually large and composed by diverse RNA-protein complexes. They typically contain numerous proteins and are highly dynamic.

At least 20 major hnRNP proteins exist in human cells, with molecular weight ranging from 34 to 120 kDa (Dreyfuss et al., 1993). Multiple isoforms of hnRNP proteins are produced by alternative processing of the genes coding for these proteins. This diversity is further increased by post-translational modification of potential physiological significance, including dimethylation of arginine side chains and phosphorylation. Most hnRNP proteins have RNA-binding ability, attributable in many cases to RNP domains or other RNA-binding motifs. The

structure of the members of this family varies considerably: (a) hnRNP I (also known as the polypyrimidine-tract-binding protein), L and M contain four RNP domains; (b) hnRNP A/B proteins contain two RNP domains at the N-terminus and a Gly-rich auxiliary domain at the carboxy end; (c) hnRNP C proteins contain a single RNP domain at the N-terminus (d) hnRNP E1 and E2 proteins contain three KH domains.

Most hnRNP proteins have general RNA-binding activity, but individual proteins display preference for specific sequences. However, hnRNP proteins generally do not bind specific sites exclusively but recognize different RNAs with a wide spectrum of affinities. Preferred sequences tend nevertheless to coincide with sites of functional importance in pre-mRNA processing, suggesting that hnRNP proteins may form specialized complexes or indirectly recruit other factors to such sites. RNA binding is further modulated by cooperative protein-protein interactions. The array of hnRNP proteins bound to a given hnRNA is determined by the RNA sequence (Dreyfuss et al., 1993).

Under conditions of competition for distinct binding sites, each mRNA associates with a unique set of hnRNP proteins, and the composition of these complexes is likely to change during the processing and transport of hnRNAs. hnRNP proteins function in early processing events by associating with nascent transcripts. hnRNP A1 modulates splice site selection by antagonizing the activity of serine-arginine-rich splicing factors (SR proteins, see below) (Caceres et al., 1994). In fact in at least some cases, hnRNP A1 cooperatively binds exonic splicing silencer sequences within the same exon and blocks binding of the SR proteins to exonic splicing enhancers (Zhu J, 2001).

Alternative splicing is a major mechanism to control the expression of viral and cellular genes, and variations in the intracellular concentration of antagonistic splicing factors may affect tissue-specific and developmental regulation of gene expression. The amount of hnRNP A1 varies during the cell cycle, but the physiological significance of this variation remains

unclear. Splicing regulation by hnRNP A1 requires both RNP domains; both domains are also required for in vitro specific RNA recognition.

SR proteins

SR proteins are a family of pre-mRNA splicing factors that share a common structure comprising one or two RNP domains at the N-terminus and a domain rich in arginine-serine dipeptides at the C-terminus (RS domain). Other proteins (for example snRNP associated factors such as U1 70K) share this structural organization. However, the term SR proteins refers to ten major polypeptides ranging from 20 to 75 kDa in size that share the ability to modulate splice site choice. SR proteins are essential pre-mRNA splicing factors and critical regulators of the selection of both constitutive and alternative splice sites. They are present in all eukaryotes except yeast, where true alternative splicing does not occur.

SR proteins have been implicated in many steps of spliceosome assembly, from recognition of the 5' splice site at the earliest step of splicing (Kohtz et al., 1994; Zuo and Manley, 1994) to binding and regulation of exonic-enhancer sequences that stimulate use of sub optimal splice sites. SR proteins differ for the presence or absence of the second RNP domain and for the length of the RS domain; when present, the sequence of the second domain is often divergent from the canonical consensus sequence (Fu, 1995; Manley and Tacke, 1996). It is apparent that sequence-specific binding to distinct pre-mRNAs is crucial for the function of these proteins in the earliest steps of spliceosome assembly (Fu, 1993; Manley and Tacke, 1996). Constitutive as well as alternative and tissue-specific splicing seem to be regulated by the different RNA-binding specificities observed for different SR proteins (Fu, 1993; Tacke et al., 1997; Tacke and Manley, 1995). The RNP domain(s) are essential for the activity of all members of this family (Chandler et al., 1997) and define the RNA-binding specificity of individual proteins (Wu and Maniatis, 1993).

Alternative splicing is a common mechanism to regulate gene expression in a cell-type specific manner, because it allows the synthesis of different proteins from the same gene. Since cell-type specific differences are observed in SR proteins distribution, it has been suggested that each cell type might have different concentrations of each SR protein (Fu, 1993). Their activity in alternative splicing is antagonized by members of the hnRNP A/B family. Thus, the counteracting activities of these two families of antagonistic factors can regulate alternative splicing both in vitro and in vivo (Caceres et al., 1994), suggesting that the physiological concentration of competing splicing factors may be important for regulation of splice site selection. The activity of SR proteins in regulation of alternative splicing is closely connected with the role of exonic enhancer sequences. In fact, these trans acting factors interact with the ESEs enhancing the selection of particular splice sites in order to promote inclusion of alternative exons with weak splice sites in the pre-mRNA (Cartegni L, 2002) probably by recruiting components of the splicing machinery to the nearby intron (Reed, 1996) (Fig. 1.5a), or by antagonizing the negative effect on splicing of an inhibitory protein that is bound to a juxtaposed exonic splicing silencer (ESS) (Fig. 1.5b).

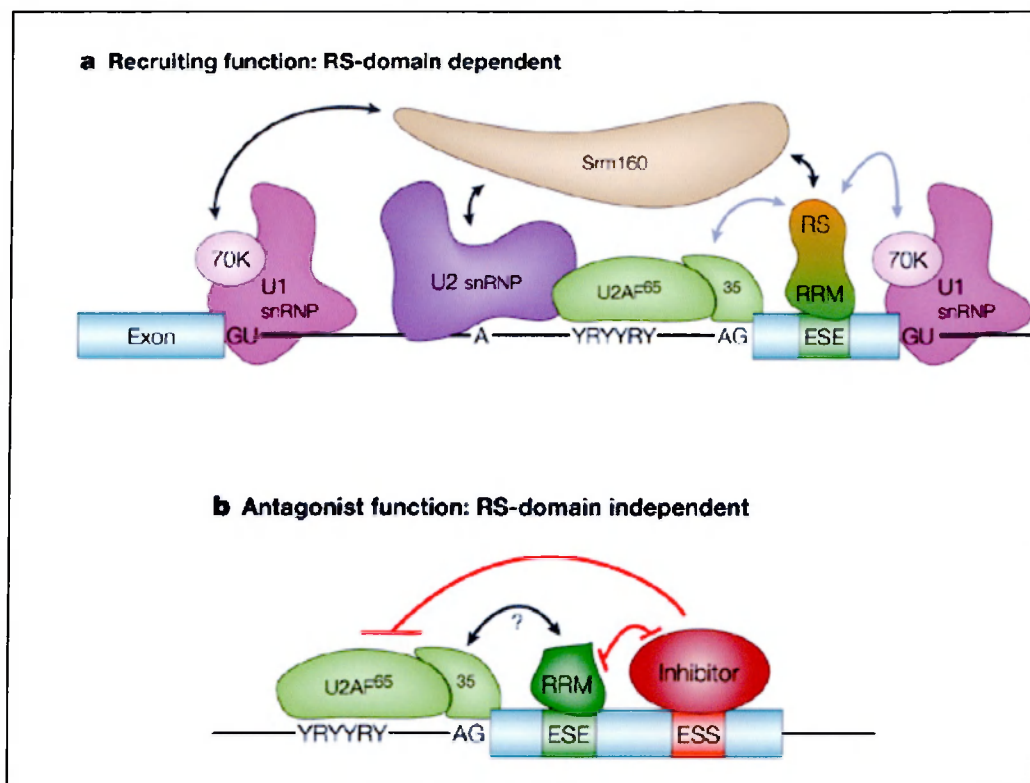


Figure 1.5 Models of SR protein action in exonic-splicing-enhancer-dependent splicing. a) RS-domain-dependent mechanism. An SR protein binds to an exonic splicing enhancer (ESE) through its RNA-recognition motifs (RRM) and contacts the splicing factor U2AF35 (U2 auxiliary factor) and/or U1-70K at the adjacent splice sites through its RS domain. The U2AF splicing factor consists of two subunits (U2AF65 and U2AF35), and the large subunit binds to the polypyrimidine (Y) tract, which here is interrupted by purines (R) and is therefore part of a weak 3' splice site. U2AF65 also promotes binding of U2 snRNP to the branch site. U2AF35 recognizes the 3' splice-site AG dinucleotide. The U1 snRNP particle binds to the upstream and downstream 5' splice sites through base pairing of the U1 snRNA; the 70K polypeptide of each U1 snRNP particle is shown. The three sets of splicing-factor–pre-mRNA interactions (U2AF–3' splice site, U1 snRNP–5' splice site and SR protein–ESE) are strengthened by the protein–protein interactions (blue arrows) that are mediated by the RS domain. For some ESE-dependent pre-mRNAs, indirect interactions (black arrows) are bridged by the splicing co-activator Srm160, which stimulates splicing of some ESE-dependent pre-mRNAs and also interacts with the U2 small nuclear ribonucleoprotein (snRNP)15. b) RS-domain-independent mechanism. Here, the main function of the SR protein that is bound to an ESE is to antagonize the negative effect on splicing of an inhibitory protein that is bound to a juxtaposed exonic splicing silencer (ESS). The SR protein is shown without its RS domain, although this domain is normally present and might still promote U2AF binding, or other domains might be involved in protein–protein interactions. Inhibitory interactions are shown (red), as is a putative stimulatory interaction (double-headed arrow). These models are not mutually exclusive, and the splicing of some introns might involve a combination of these mechanisms Figure adapted from Cartegni et al. (Cartegni L, 2002)

The structural organization of SR proteins and their ability to interact with other proteins suggest a model for their function, where the RNA-binding domain(s) provide specificity for certain mRNAs (or classes of mRNAs) by recognizing specific RNA signals, while the RS domain may recruit other factors during spliceosome assembly by protein-protein interactions (Valcarcel and Green, 1996)

Activity of SR proteins is regulated through phosphorylation/dephosphorylation cycles. In fact, SR proteins are phosphorylated *in vivo* at multiple serine residues within the RS domain (Colwill et al., 1996a; Colwill et al., 1996b; Gui et al., 1994). Phosphorylation and dephosphorylation seem to be crucial to define the organization of splicing inside the cell nucleus by affecting the RNA-binding activity and sub nuclear localization of RS-domain containing proteins (Misteli et al., 1997). Phosphorylation is important for specific RNA recognition, since the high positive charge of unphosphorylated RS domains masks the specificity of the RNP domains and enhances non-specific binding (Tacke et al., 1997; Xiao and Manley, 1997). Protein-protein interactions are enhanced by phosphorylation (Xiao and Manley, 1997), which also affects the subnuclear localization of splicing factors, causing their release from storage sites of splicing components (Colwill et al., 1996b; Xiao and Manley, 1997).

Splicing-dependent exon-exon junction complex

Recent studies have shown that several proteins become associated with spliced mRNAs that acquire and/or lose factors in a splicing-dependent, sequence-independent manner (Le Hir et al., 2000b). *In vitro* splicing experiments showed that mRNP complexes produced by splicing have a different mobility on a non-denaturing gel compared to mRNP complexes containing intronless mRNAs with identical sequences (Luo and Reed, 1999). It has been shown that mRNA is bound to a complex of proteins that is deposited at a site of 20-24 nt upstream of an

exon-exon junction (Kataoka et al., 2000; Le Hir et al., 2000a; Luo and Reed, 1999). This “exon junction complex” (EJC) can include at least 7 proteins, although not all of them may be present simultaneously: RNPS1 (Mayeda et al., 1999), SRm160 (Blencowe et al., 1998), DEK (McGarvey et al., 2000), TAP (Bear et al., 1999), Aly/REF (Zhou et al., 2000), Y14 (Kataoka et al., 2000), Upf3 (Kim et al., 2001) and Mago (Le Hir et al., 2001).

Human RNPS1 was shown to interact with SR proteins and strongly activate splicing of both constitutively and alternatively spliced pre-mRNAs (Mayeda et al., 1999). DEK co-purifies with SR proteins as well and can interact with SRm160 in vitro (McGarvey et al., 2000). However, the specificity of its association with spliced mRNA is less than that of other components. SRm160 enhances splicing of specific pre-mRNAs through interactions with snRNPs and other SR proteins (Le Hir et al., 2000a). Several lines of evidence suggest that TAP play a major role in cellular mRNA export pathway. In fact, TAP shuttles from the nucleus to the cytoplasm (Bear et al., 1999). Then, it interacts with nucleoporins in vitro and localizes to nuclear pores as well as to the nucleoplasm (Katahira et al., 1999). Finally, tethering TAP to unspliced mRNA enhanced the export, especially in the presence of the cofactor, p15/NXT1 (Guzik et al., 2001). Other RBD-containing proteins, Y14 and Aly/REF, also bind efficiently and preferentially to spliced mRNAs. Y14 and Aly/REF shuttle between the nucleus and the cytoplasm suggesting that these proteins play roles in the export of mRNA to the cytoplasm and/or in cytoplasmic processes. In fact, Aly/REF was found to stimulate mRNA export (Rodrigues et al., 2001; Zhou et al., 2000). Mammalian Aly/REF can interact directly with TAP and RNA at the same time in vitro suggesting that Aly/REF recruits TAP to mRNA.

Recent studies provide direct evidence that hUpf3 is part of the splicing-dependent exon-exon junction complex (Kim et al., 2001; Lykke-Andersen et al., 2001). hUpf3 associates with spliced mRNAs at -20nt relative to exon-exon junctions and interacts with specific

components of the complex, Y14 and RNPS1. Furthermore, tethering of RNPS1 or Y14 to the 3' untranslated region triggered nonsense mediated decay of the mRNA. Thus the splicing-dependent, position-specific binding of hUpf3 to mRNAs through interactions with RNPS1 and/or Y14 likely provides the long-sought link between splicing and the NMD pathway.

Magoh protein (human homolog of *Drosophila mago*), binds avidly and directly to Y14 and TAP, but not to other known components of the complex, and is found in Y14-containing mRNPs in vivo. Importantly, Magoh also binds to mRNAs produced by splicing upstream (approximately 20 nucleotides) of exon-exon junctions and its binding to mRNA persists after export. These experiments thus reveal specific protein-protein interactions among the proteins of the splicing-dependent mRNP complex and suggest an important role for the highly evolutionarily conserved magoh protein in this complex.

The functions of SRm160, DEK, RNPS1, and Y14 in post-splicing events are not yet clear. These spliced mRNA-associated proteins were found to form a stable complex at approximately -20nt relative to exon-exon junctions in vitro and in *Xenopus* oocytes (Le Hir et al., 2000b) (Fig. 1.6). The complex appears to be dynamic since Aly/REF rapidly dissociates during export (or immediately after export) while Y14 persists in the same position on mRNAs in the cytoplasm (Kim and Dreyfus, 2001). The size of the complex was estimated to be 335 kDa in vitro and the shortest RNA fragment that is bound to the complex was of 8 nucleotides in length (Le Hir et al., 2000b). However, it remains to be determined whether this complex is assembled at every exon-exon junction of the mRNA or only at some specific junctions. Apparently, the variety of components that form this complex makes it likely that it can be involved in several processes. Two functions in particular, mRNA export and nonsense-mediated decay, are likely functions for the splicing-dependent exon-exon junction complex.

Splice junction pairing: Exon definition and Intron definition

Although the snRNPs are some of the responsible for locating intron-exon junctions by base pairing, this alone does not explain how authentic splice sites are recognized when sequences that match the consensus splice sites are likely to occur elsewhere in the pre-mRNA; and how the splicing machinery joins the exons in the correct order. How the snRNPs and splicing proteins recognize the proper sequences is largely unknown. How splicing occurs in the correct order has been explained. Some studies have shown that splicing factors bind to RNA polymerase II and proposed that the factors assemble onto the pre-mRNA as the RNA emerges from the transcription complex (Du and Warren, 1997; Misteli and Spector, 1999; Yuryev et al., 1996). Presumably, this helps distinguish introns and exons and so an snRNP at a given 5' splice site will see only the next emerging 3' splice site because sites further along the transcript have not yet been synthesized. This might also prevent exon skipping.

Regarding the correct joining of exons, in 1990 Susan Berget proposed a model based on pairing between the splice sites across an exon (Robberson et al., 1990). This model stems from the observations that in human and, more in general in vertebrates a typical primary transcript can be until 30 kilobases long and contains about ten exons separated by much larger and more variably sized introns. The discrepancy between the length of human exons in comparison with that of introns (Hawkins, 1988) led to the "exon definition" model of splicing in which splice sites are first paired across exons, with subsequent spliceosome assembly proceeding through pairing of exon units. In the alternative "intron definition" model, splice sites are initially paired across introns rather than exons (Fig.1.6).

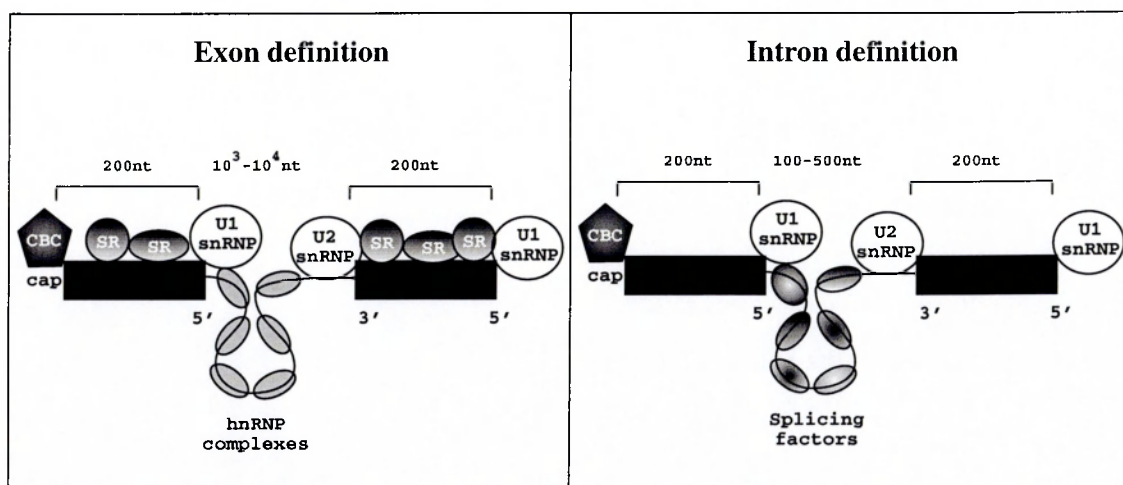


Figure 1.6 Exon definition versus Intron definition. In the “Exon definition” model the exons are recognized as a unit during early spliceosome assembly. Multiple factors interact with exonic sequences defining the 5’ and the 3’ splice site of the exons. In the “Intron definition” mechanism the introns are recognized as a unit. Multiple factors favour the 5’ and the 3’ splice site pairing of the introns.

Intron definition is thought to be the predominant mode of splicing in transcripts containing short introns and long exons. Several lines of evidence support the “exon definition” model. Moreover, it is apparent that both exon definition and intron definition might occur in different parts of the pre-mRNA of the same gene.

Evidence for Berget's “exon definition” model arises from the observation that the first and the last exon should require a special mechanism for their recognition. It was shown that the 5’ capping and the 5’ splice site are necessary to define the first exon (Izaurralde et al., 1994). On the other hand, last exon is defined by 3’ splice site and polyadenylation signal (Gunderson et al., 1994). Moreover, it was shown that exon length might affect splicing. In fact, a minimal separation of the sites seems to be required to prevent steric hindrance of the factors that recognize the splice sites bordering an exon. This indication is supported by the observation that a constitutively recognized internal exon was skipped by in vivo splicing machinery if its size was smaller than 50 nucleotides (Dominski and Krole, 1991). In addition, Black et al. (Black, 1991) have shown that extension of the 18-nucleotide N1 exon of the mouse c-src

gene up to 109 nucleotides leads to its constitutive inclusion. This finding suggested that the exon is normally skipped because it is too short to allow spliceosomes to assemble at both ends simultaneously. On the other hand, the expansion *in vitro* of internal exons to length above 300 nucleotides leads to activation of the cryptic splice site inside the exon or to exon skipping (Berget, 1995) indicating that splicing efficiency is affected by length. In fact, less than 1% of the known internal exons in vertebrate are longer than 400 nucleotides. Additional support for the “exon definition” model derives from *in vitro* (Kuo et al., 1991; Talerico and Berget, 1990) and *in vivo* experiments (Dominski and Kole, 1991; Dominski and Kole, 1994; Xu et al., 1993). These studies revealed how the strength of splice sites can affect the splicing of both the introns flanking the exon and not only of the intron bearing the mutated splice site, as predicted from the “intron definition model”. The intron definition model suggests that the intron is the unit recognized by the splicing machinery and proposes a scanning mechanism where the 5' and 3' splice signals (5'ss and 3'ss) are initially recognized and paired across the intron. (Lang and Spritz, 1983). Evidence supporting such a mechanism arises from observations made in yeast, where messengers often have unique introns and their length is usually below 100 nucleotides (Goguel and Rosbash, 1993) and *Drosophila* where most exons are 100 to 180 nucleotides in length (Hawkins, 1988). In vertebrate most exons are small and are flanked by long introns. However, some exceptionally long exon exists and when an expanded exon is placed in a gene in which the flanking introns are small, the exon is constitutively included. Experiments on a three exons minigene, containing short intronic sequences, showed that an expanded size of the middle exon up to 1400 nucleotides does not lead to exon skipping (Chen and Chasin, 1994).

Alternative splicing

Alternative RNA splicing is the process that allows the selection of different combination of splice sites within precursor mRNA. This process is seen in nearly all metazoan organisms as a means for producing functionally diverse polypeptides from a single gene (Lopez, 1998).

Alternative splicing is especially common in vertebrates. Alignment of EST sequences and mapping the resulting mRNA families to the human genome provided a minimum estimate that 35% of human genes show variably spliced products (Croft et al., 2000). However, since ESTs derive from a limited number of tissues or developmental stages, and cover a little portion of each mRNA, the true percentage is likely much higher. There are also remarkable examples of hundreds and even thousands of functionally divergent mRNAs and protein being produced from a single gene. A recent discovery in *Drosophila* is a fascinating example of the extremely high number of proteins that can be produced from one gene. The DSCAM gene presents different exons used in a mutually exclusive manner. For each of these exons there are different numbers of alternative forms. If all the combinations of these exons were used, the single DSCAM gene would produce 38,016 different DSCAM proteins (Schmucker et al., 2000). Variation in mRNA structures takes many different forms (Lopez, 1998). Exons can be spliced into the mRNA or skipped. Introns that are normally excised can be retained in the mRNA. The position of either 5' or 3' splice sites can shift to make exons longer or shorter.

In addition to these changes in splicing, alterations in transcriptional start site or polyadenylation site also allow production of multiple mRNAs. All of these changes in mRNA structure can be regulated in diverse ways, depending on sexual genotype, cellular differentiation, or the activation of particular cell signaling pathways.

The mechanisms that determine which splice site are utilized and how this selection is regulated in different cell types or developmental stages have still not been precisely defined.

Much progress has been made in identifying the cis-acting elements and the trans-acting factors involved in the alternative splicing.

The human Apolipoprotein AII exon 3 as a model to study the influence of atypical polypyrimidine tract (GU) on splicing

The Apolipoprotein AII gene

The human apolipoprotein AII gene (ApoAII) encodes apolipoprotein AII, a major constituent of the high-density lipoprotein (HDL) particles that play important roles in reverse cholesterol transport and lipid metabolism. Transcription of ApoAII occurs primarily in the liver (Hussain, 1990) and is controlled by a complex array of positive and negative regulatory elements located -911 to +29 bp relative to the cap site of the gene (Chambaz et al., 1991). The position of known transcription-factor-binding sites, relative to the exons of the gene, is shown in Figure 1.7. Very few genetic variants, either at the ApoAII gene itself, or in its upstream promoter, have been identified. To date, no protein-level polymorphism of ApoAII has been reported (Civeira et al., 1992). Furthermore, only one rare variant, a splice-junction mutation at the start of the third intron of the gene that results in complete lack of ApoAII expression in two individuals from a single family (Deeb et al., 1990), has ever been identified; however, the ApoAII deficiency associated with this mutation has no effect on HDL cholesterol levels.

At the DNA level, three polymorphisms have been described previously: a multiallelic (GT) microsatellite in the second intron of the gene (Weber and May, 1989) (position 1671 in Fig. 1.7); a C->T single nucleotide polymorphism located 528 bp 3' of the gene within an Alu repeat sequence (Civeira et al., 1992) (position 2994 in Fig. 1.7) and identified originally as a

MspI restriction fragment length polymorphism (Scott et al., 1985); and a C->T SNP located in the third intron (Dupuy-Gorce et al., 1996) (position 2233 in Fig. 1.7).

Interestingly, microsatellite sequences, such as (GT)_n repeats, constitute probably one of the best source of multiallelic microsatellite-polymorphisms in the human genome (Tautz and Renz, 1984; Weber and May, 1989). Although the function of these microsatellites is unknown, it is known that their mutation rate is very high and it has been suggested that they can operate as hot spots for recombination (Hellman et al., 1988) or be involved in gene regulation (Wittig et al., 1992).

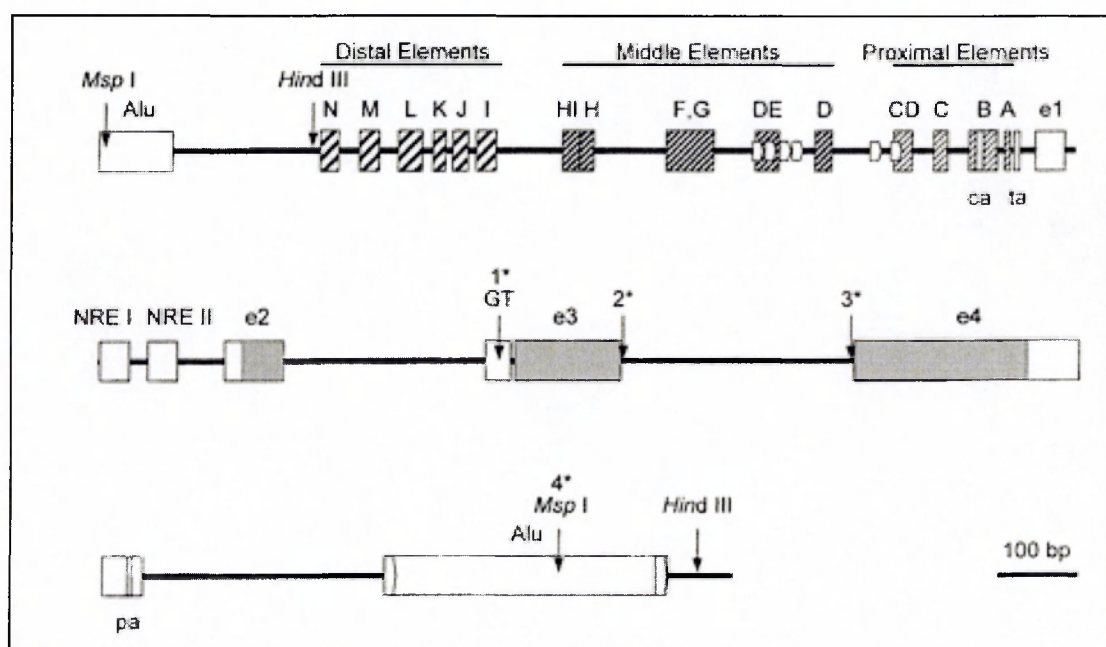


Figure 1.7. Schematic representation of the *apoA-II* gene. Several regulatory elements are located in the promoter region (-911 to +29), which can be divided into three functional regions. Arrows indicate the presence of direct repeats in the promoter region. Also, the first intron appears to contain negative regulatory elements (NRE I and II). ca, CAAT box; ta, TATA box; pa, polyadenylation signal. Polymorphic sites are indicated by an asterisk: 1*, GT repeat region; 2*, G-to-A transition that affects the donor splice site of third intron; 3*, C-to-T transition near the acceptor splice site of intron 3 that also modifies a *Bst*NI restriction site; 4*, C-to-T transition that modifies an *Msp*I restriction site within the 3' Alu element. Figure adapted from Blanco-Vaca et al (2001)

Polymorphic (GT)_n within the 3' splice site of human ApoAII intron 2

Polymorphic (GT) microsatellites have been found in different human genes close to the 3'splice site of some exons. Cardiac muscle actin intron 4, Apolipoprotein AII intron 2, and CFTR intron 8, share the peculiar feature of including these polymorphic microsatellites within the polypyrimidine tracts (Fig. 1.8).

	3' splice site	5' splice site
c- α -actin	(t g)25 a c t c g t t c c c a g / G T A T G G...A T T	A A G / g t a a a g a a
Apo AII	(t g)m g g g c a g / G A G C T T...G G C	C A A / g t a a g t c t
hCFTR	(t g)m t n a a c a g / G G A T T T...G G C	A A G / g t a g t t c t
mCFTR	t c t t c a a t g a a t a g / G G A T T T...G G C	C A A / g t a a g t c t
Consensus	(Y)n n c a g / G T ...	C A G / g t a a g t a c

Figure 1.8. Comparison of splice sites from the human Apo AII exon 3, human CFTR exon 9, human α cardiac actin and mouse CFTR exon 9 with consensus sequences for 5' and 3' splice sites. 3'- and 5'-splice sites of the different genes (upper line) are compared with consensus sequences (lower line). Y pyrimidine. Upper case, exons; lower case, introns.

On one hand, in the intron 8 of the CFTR gene, there are two polymorphic elements juxtaposed each other, a polythymidine tract and immediately upstream a GT repeats. Both are adjacent to the splice junction intron 8/exon9. The number of GT repeats ranges from 9 to 13 and their distance from the 3'splice site is variable, due to the presence of the T tract also polymorphic for its length.

A population study suggested the negative influence of GT stretch on correct splicing of exon 9 by analyzing the genotype of patients affected by congenital bilateral absence of vas deferens (CBAVD). The patients were found to carry a high number of GT repeats, while healthy subjects harbored a low number of GT repeats. The conclusions of the studies carried out by Cuppens et al. was that a longer GT repeats would place the branch point A nucleotide of the lariat in a less favorable position for splicing (Cuppens et al., 1998). In fact, the putative

branch points of CFTR intron 8 is placed at about -40 bases upstream of the 3' splice site depending on the length of the GT/T tracts.

Then, *in vivo* studies demonstrated the relevance of the GT repeats for the splicing of CFTR exon 9 when they are located within the polypyrimidine tract region. Niksic et al showed that the increase in the length of the GT repeats within the 3' splice site of intron 8 is directly correlated with a detrimental effect in the inclusion of exon 9 in pre-mRNA (Niksic et al., 1999). In addition, it was observed that the deletion of the T tract resulted in 100% of exon 9 skipping, and this finding strongly indicated that the GT tract alone is not able to work as a functional polypyrimidine tract in the CFTR intron 8 context.

On the other hand, the second intron of the human ApoAII also presents a polymorphic GT tract in the polypyrimidine region. In this case 10 different alleles of the ApoA-II (GT)_n repeat polymorphism have been identified, corresponding to 12–21 repeats. Allelic distributions show that four alleles [(GT)₁₃, (GT)₁₄, (GT)₁₆ and (GT)₁₉] account for 94.6% of the total. The most frequent allele is the (GT)₁₉ (33.4%), the second most common is the (GT)₁₆ allele (28.9%), followed by the (GT)₁₃ (17.4%), and the (GT)₁₄ alleles (14.9%) (Brousseau et al., 2002).

Association studies have shown a significant correlation of the (GT)₁₉ allele with a decrease of plasma ApoAII concentration, whereas no correlation with the alleles (GT)₁₃, (GT)₁₄, (GT)₁₆ and ApoA-II plasma concentration has been found. In fact, the (CA)₁₉ allele was associated with a decrease in total cholesterol as well as significantly decreased ApoA-II and LpA-II:A-I plasma levels. (Brousseau et al., 2002).

Then, *in vivo* studies have indicated that the GT repeats within ApoAII intron 2 are important for exon 3 splicing (Shelley et al., 1985). In fact, these studies have shown not only that this (GT)_n polymorphisms located near the acceptor site of intron 2 works efficiently as a polypyrimidine tract when n=16 (Shelley et al., 1985), but also that the six nucleotides

(GT)₂GG, which directly replace this tract in a deletion mutant of the 3' splice site of ApoA-II intron 2, are sufficient to direct the splicing process efficiently and correctly (Shelley and Baralle, 1987).

The use of a poly GT repeats as a functional polypyrimidine tract has been also studied by in vitro experiments where the GT repeats works in different contexts (Coolidge et al., 1997).

Lastly, 12 different (GT)_n microsatellite alleles have been detected within the intron 4 of the human cardiac actin gene (Litt and Luty, 1989). Although no studies have been carried out in order to study the possible effects of this polymorphic GT repeats with transcription or splicing process of cardiac actin exon 5, so far no alternative splicing of this exon has been described.

Considering all these studies, it is evident that the apparent similarity of the polypyrimidine tract, concerning the GT repeats, is contrasted by the different splicing pattern exhibited by CFTR exon 9, ApoAII exon 3 and cardiac actin exon 5. In fact, CFTR exon 9 undergoes alternative splicing to a variable extent, whereas ApoA-II exon 3 and cardiac actin exon 5 are always constitutively included in the mRNA.

The alternated purine-pyrimidine tracts are known for their propensity to adopt a Z-DNA conformation that show to play an active role in transcription. Since there are several lines of evidence that in eukaryotes the splicing can occur co-transcriptionally (Eick, 1990), the presence of a Z-DNA forming region might interfere with the splicing process.

AIM of the RESEARCH

The 3' splice site of the second intron of the human Apolipoprotein AII gene is unique because a stretch of alternating guanine and thymine residues, (GT)₁₆GGGCAG, replaces the polypyrimidine tract usually associated with 3' splice junctions. Previous studies have demonstrated that it is important for the constitutive splicing of ApoAII exon 3. In fact, transient transfections of progressive 5' deletion mutants have shown that (GT)₂ is the minimum number of repeats at the 3' end of Apo AII intron 2 that are required to direct efficient exon 3 splicing. When all the (GT) repeats were deleted, the 3' splice site of ApoAII intron 2 became completely non-functional and exon 3 was totally skipped.

Interestingly, the human Apolipoprotein AII intron 2 shares the peculiar architecture of the 3' splice site based on a stretch of alternating guanine and thymine residues, with the eighth intron of human CFTR gene. The polypyrimidine tract of human CFTR intron 8 is composed by a (GT)₉₋₁₃ tract followed by a polythymidine tract.

In spite of the presence of a (GT) tract within the polypyrimidine tract of the Apo AII intron 2 and the human CFTR intron 8, it is evident the striking difference in the splicing behavior of the downstream exons: human Apo AII exon 3 is constitutively spliced whereas the human CFTR exon 9 undergoes alternative splicing.

These observations suggest the possible presence of accessory cis-elements within the Apo AII exon 3 and/or the flanking introns that can support its constitutive inclusion in ApoAII pre-mRNA.

The aim of the present thesis was to investigate the mechanisms underlying the constitutive splicing of ApoAII exon 3 and, in particular, to characterize the cis-acting elements and the trans-acting factors involved in ApoAII exon 3 definition.

2. RESULTS

Analysis of the ApoAII exon 3 splice sites

Current consensus indicates that exon definition is mediated at least in part by SR proteins bridging between the small nuclear ribonucleoprotein (snRNP) U1 at a downstream 5' splice site and the U2AF heterodimer at the upstream 3' splice site. It has been previously mentioned that CFTR exon 9 and ApoAII exon 3 share the almost unique characteristic of a polymorphic (GU) tract at or close to the 3' AG dinucleotide. However, the influence of the (GU) tract on splicing efficiency seems to be different. In fact, for the CFTR exon 9, the longer is the (GU) tract the higher levels of exon 9 exclusion are observed. In the ApoAII exon 3, variations in the length of (GU) tract do not seem to affect the splicing efficiency of exon 3 that is constitutively included in pre-mRNA. These observations suggest that the context of these two genes may differ in critical elements affecting splicing. The more obvious difference is the short U tract in the CFTR 3' splice site but it cannot explain the dramatic different behavior.

As a first approach, we wished to examine whether an association exists between the different splicing behavior of human ApoAII exon 3, human CFTR exon 9 and splice site strength.

Therefore, we compared the base sequence composition of the 5' and 3' splice sites of the human ApoAII exon 3, human CFTR exon 9 and considered also mouse ApoAII exon 3 that is constitutively spliced likewise human ApoAII exon 3 but has a more classical polypyrimidine tract. The strength of these splice sites were calculated according to the Splice Site Prediction by Neural Network program (SSPNN, http://www.fruitfly.org/seq_tools/splice.html).

Human and mouse 5' splice junctions are identical until base +8 of intron 3. Both deviate from the consensus sequence at position -1, where the consensus base G is replaced by another purine, A, and in position +7 and +8, where the consensus bases A and C are replaced by C and T, respectively (Fig. 2.1).

The analysis of the 3' splice junction shows that human and mouse have a good match with the consensus even if they contain a different pyrimidine (C in human and T in mouse) in position-3 (Fig. 2.1). The human ApoAII exon 3 "polypyrimidine tract" is composed of (GU)_n repeats that range in different individuals between 16 and 19 being 16 the more frequent allele. Conversely, the mouse polypyrimidine tract does not present (GU) repeats.

	ACCEPTOR	DONOR
mAII	ucauugcuggcagucuaacaugcagaggcugugaucuauguuuguguag/GAGCUUU...GGCCAA/guaagucu	
hAII	gcugaaaagagugugugugugugugugugugugugugugugugugggcag/GAGCUUU...GGCCAA/guaagucu	
hCFTR	uuuuuuuuuugauguguguguguguguguguguguguguuuuuuuacag/GGAUUUG...GGCAAG/guaguucu	
Consensus	(Y)_n yag/GU YAG/guraguac	

Figure 2.1. Comparison of splice sites from the mouse and human ApoAII exon 3 with human CFTR exon 9. Acceptor and donor splice sites (centre) of mouse (mAII), human ApoAII exon 3 (hAII) and human CFTR exon 9 (hCFTR) aligned are compared with consensus sequences (bold lower line). Y pyrimidine and R purine. upper case, exons; lower case, introns.

The comparison of the 5' splice junction of CFTR exon 9 with the consensus sequence shows that it differs in position +4, where consensus base A is replaced by another purine, G, and in position +5, where the consensus purine G is replaced by a pyrimidine T. Therefore, the 5' splice sites of both human and mouse ApoAII exon 3 are closer to the consensus than the 5' splice site of human CFTR exon 9. Regarding the analysis of the 3' splice site of CFTR exon 9, it presents a good match with the consensus, but the region between the branchpoint and the 3' splice junction contains two polymorphic stretches: a polythymidine tract presenting four allelic variants in the population (with 3, 5, 7 or 9 thymidines) and, immediately upstream of the poly-T, another polymorphic tract containing an alternating pyrimidine/purine (GT) dinucleotide repetition with 5 allelic variants (with 9, 10, 11, 12, or 13 repeats).

The human CFTR exon 9 polypyrimidine tract seems to be stronger than that of both human and mouse ApoAII exon 3 because of the presence of a poly-T tract.

After comparison of the splice sites of these three exons, we calculated their strength using the Splice Site Prediction by Neural Network Program (SSPNN). In general, sequences that have a higher score (maximum 1.0 that is the consensus sequence) are considered to be stronger.

The program was able to find the authentic 5' and 3' splice sites of human CFTR exon 9 whose scores are 0.83 and 0.91, respectively (Fig. 2.2).

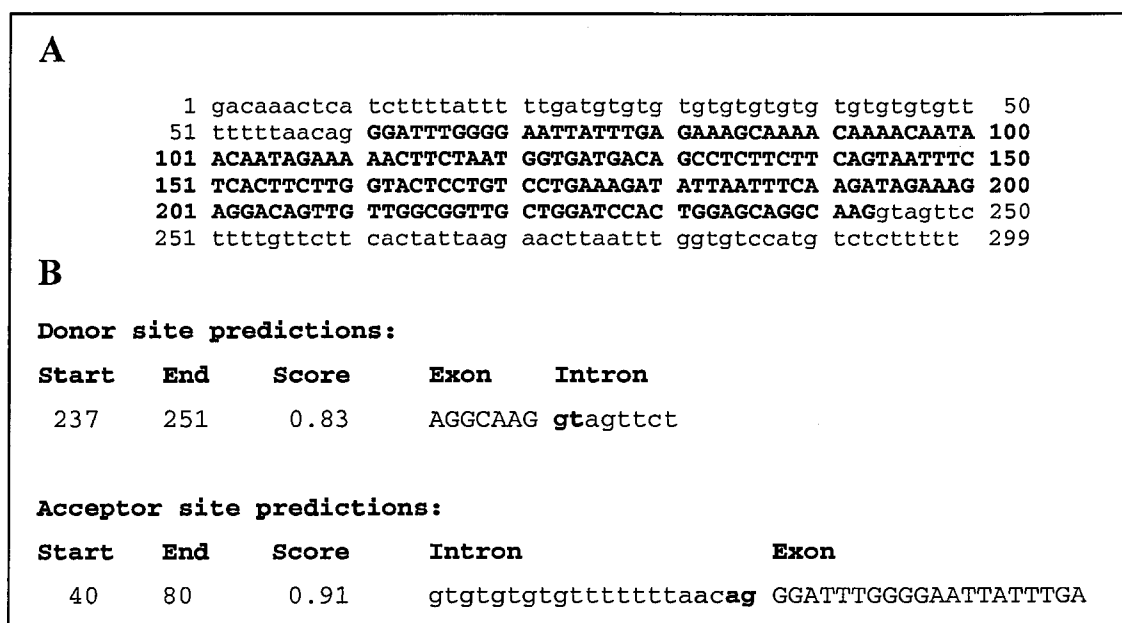


Figure 2.2. Human CFTR exon 9 splice site prediction by Neural Network. A) The 300 bp sequence encompassing exon 9 (upper case bold) and its flanking intronic regions was analyzed using the Splice Site Prediction by Neural Network program. B) Results of splice site predictions with 0.1 as minimum score for both donor and acceptor splice sites.

In the mouse ApoAII exon 3, the program identified the authentic 5' splice site (with the maximum possible score, i.e. 1.0) (Figure 2.3-B). In addition six possible 3' splice sites were also highlighted in a score range included between 0.57 and 0.99, being the score of the functional one 0.76.

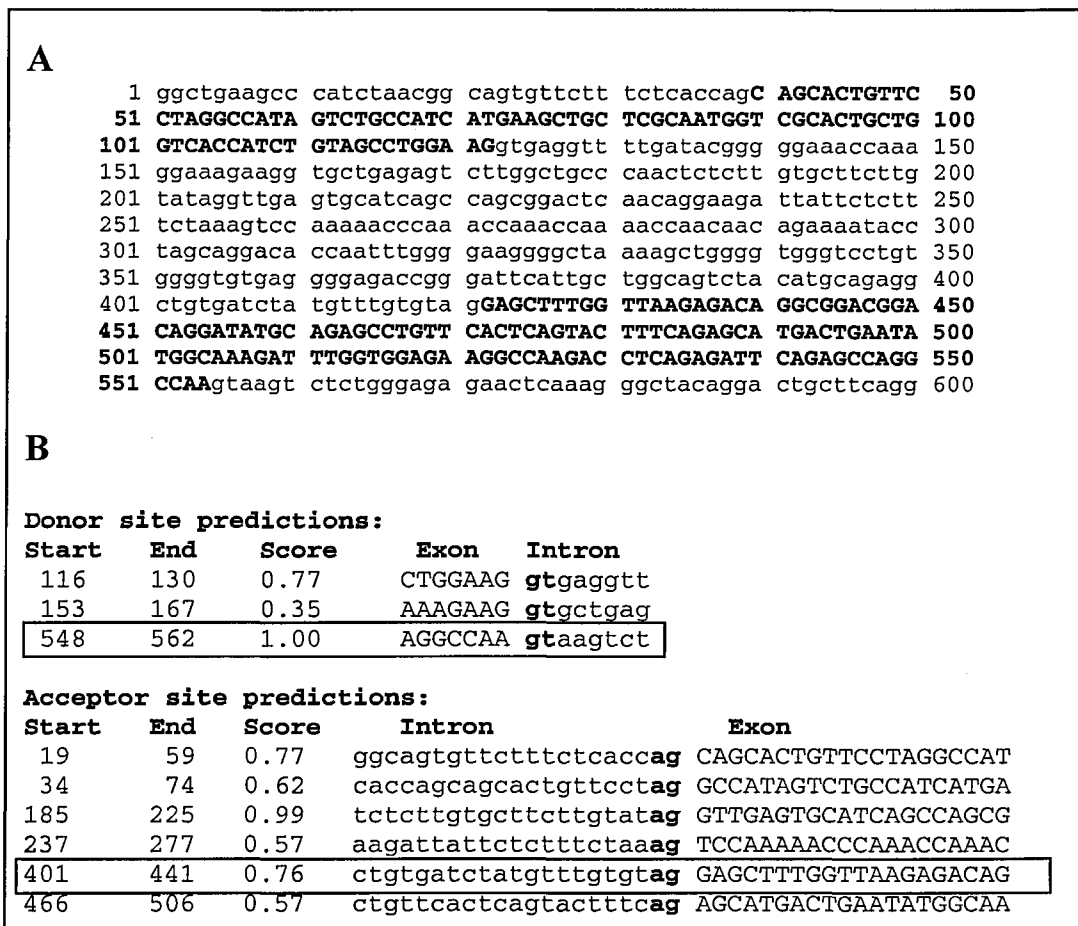


Figure 2.3. Mouse ApoAII exon 3 splice site prediction by Neural Network. A) The 600bp sequence encompassing exon 1, intron 1 and exon 2 was analyzed using the Splice Site Prediction by Neural Network program. B) Results of splice site predictions with 0.1 as minimum score for both donor and acceptor splice sites. The authentic donor and acceptor splice sites are boxed.

The calculated scores for the human ApoAII exon 3 demonstrate that the authentic donor splice site is ranked as the best possible 5' splice site (score=1.0) within the 600bp encompassing the genomic region between ApoAII exon 2 and exon 3. On the other hand the authentic 3' splice site of ApoAII exon 3 is not ranked among the possible 3' splice sites within the same genomic region (at least with both donor and acceptor cutoff 0.1) whilst five cryptic 3' splice sites (whose scores ranged from 0.11 up to 0.99) are also identified, (Figure 2.4).

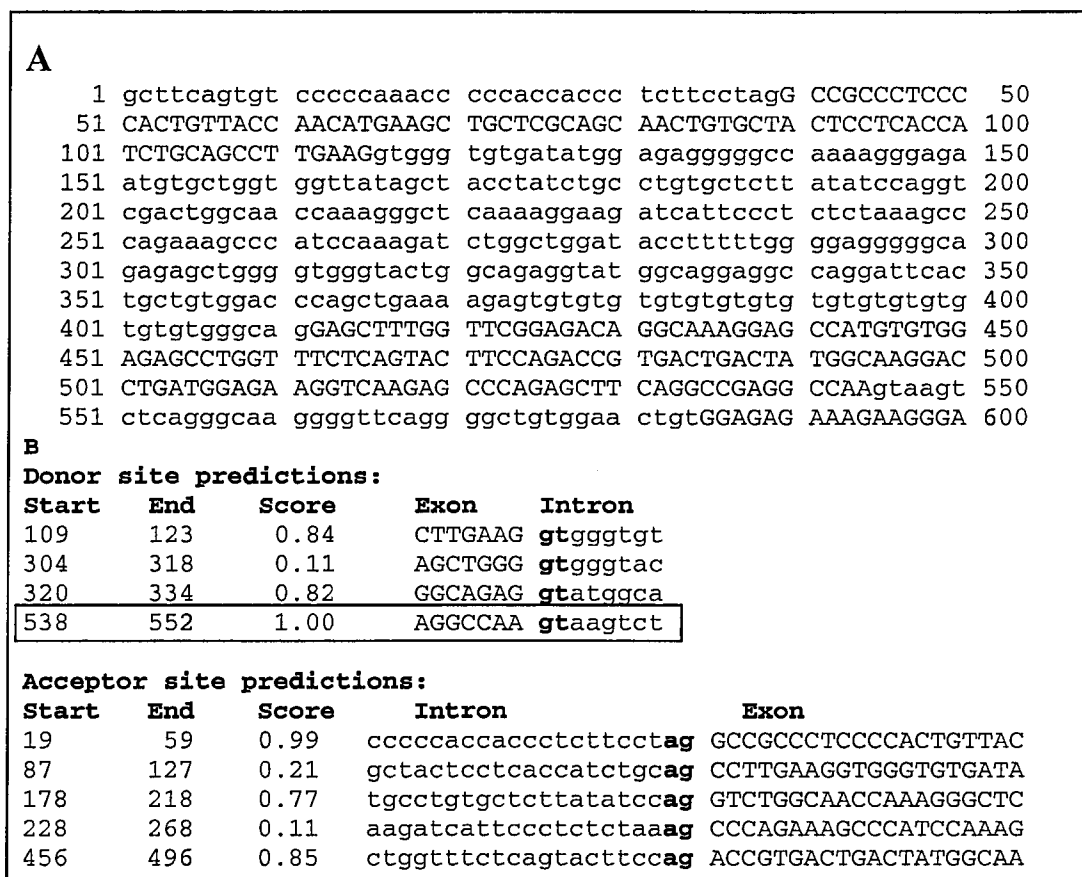


Figure 2.4. Human ApoAII exon 3 splice site prediction by Neural Network. A) The 600bp sequence encompassing exon 1, intron 1 and exon 2 was analyzed using the Splice Site Prediction by Neural Network program. B) Results of splice site predictions with 0.1 as minimum score for both donor and acceptor splice sites. The authentic donor splice sites is boxed.

Overall these observations predict that the authentic acceptor splice site of human ApoAII exon 3 is very weak because of the peculiar polypyrimidine tract structure based just on (GT) dinucleotide repeats.

The apparent weakness of human ApoAII exon 3 acceptor splice site (strength <0.1) in comparison with the human CFTR exon 9 acceptor splice site (strength =0.91) is in striking contrast with the splicing behavior of the two exons. In fact, human ApoAII exon 3 is constitutively included in mRNA whilst human CFTR exon 9 is alternatively spliced.

Hence, it is plausible that novel accessory cis-acting elements (not identified by current programs for 3' splice site analysis) might be involved in ApoAII exon 3 definition to assure its constitutive splicing.

“in vivo” and “in vitro” systems for the study of ApoAII exon 3 splicing

In order to study the relevance of the splice site strength and map the possible cis acting elements involved in the ApoAII exon 3 definition, we used both “in vivo” and “in vitro” splicing systems. For the “in vivo” approach, we generated an eukaryotic gene expression system by cloning the whole 3.2 kb ApoAII gene from the promoter to its polyadenylation site (pApo-WT, Figure 2.5).

In comparison with the minigene system widely used to study other splicing models (Kwok et al., 2003; Muro et al., 1999; Shin et al., 2003; Webb et al., 2003) where only one exon and its flanking regions are cloned in an heterologous gene context (minigene), the ApoAII expression system that we generated contains all the elements necessary for its transcription and RNA processing that are present in the endogenous ApoAII gene. Such a construct should allow the study of the cis-acting elements affecting ApoAII exon 3 splicing in a context as close as possible as the proper genomic ApoAII context.

Suitable restriction sites within intron II and intron III of the cloned ApoAII gene were included in order to generate an interchangeable cassette to study the regulatory sequences encompassing this region.

We have chosen to use the Hep3B human liver cell line in these studies because the endogenous ApoAII gene is expressed in the liver (Hussain, 1990), and hence this cell line provides an environment as close as possible as the natural one. Previous studies have demonstrated that ApoAII is expressed exclusively in liver and intestine derived cells. The expression is directed by a DNA sequences located in a 259 bp region centred 782 bp upstream from the transcription initiation site which is absolutely required for tissue specific transcription from the Apo AII promoter (Shelley and Baralle, 1987b). A target sequence was included in the ApoAII construct at the end of exon 4, to allow the specific amplification of pApo-wt derived RNA with specific primers (see materials and methods) that differentiate

between the transfected and the endogenously expressed ApoAII gene in the human hepatocarcinoma (Hep3B) cell line.

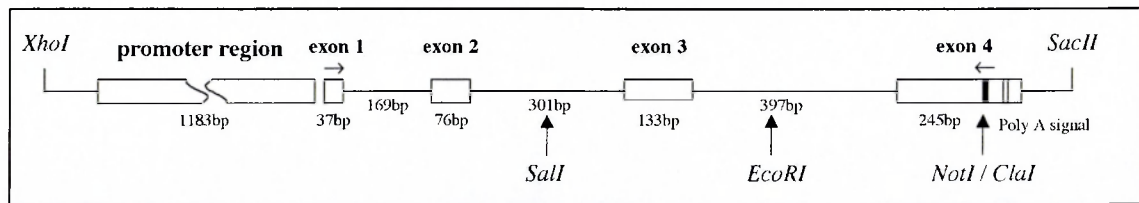


Figure 2.5. Schematic representation of pApo construct. The whole Apo AII gene from the promoter region to the poly A signal was cloned into XhoI-SacII sites of pBluescript plasmid. SalI and EcoRI sites were included in IVS2 and IVS3 respectively. Size of every exon (white boxes) and introns (solid lines) as well as position of relevant restriction enzymes target sites are shown. Primers used for RTPCR are shown as superimposed arrows annealing in exon 1 and exon 4.

The splicing pattern of the pApo-wt gene system was assessed by transient transfection of Hep3B cells. 48h after the transfection the cells were harvested, the RNA was extracted and treated with DNase to free RNA preparations from residual plasmid. After reverse transcription using a primer specific for the pApo-wt construct or the endogenous gene, the splicing pattern was then determined by PCR amplification.

The splicing efficiency of the ApoAII construct in Hep3B cells was similar to that observed for the endogenous gene (Fig. 2.6). In fact, the splicing pattern of the pApo-wt gene system showed to overlap that of the endogenous ApoAII gene, displaying a nearly 95% of exon 3 inclusion (Fig. 2.6). An identical strategy was used in all the following experiments using mutated versions of the pApo-wt gene construct.

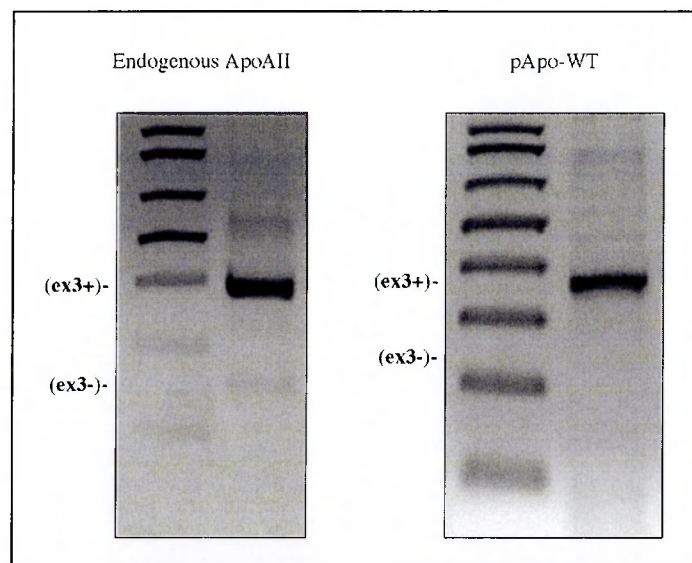


Figure 2.6. Endogenous Apolipoprotein AII and pApo-wt construct splicing pattern. Analysis of pre-mRNA splicing of endogenous Apolipoprotein AII gene and pApo construct in Hep 3B cell line. The size of PCR products including (365bp) and excluding (232bp) exon 3 are indicated. Amplicons were separated on 2.0% (w/v) agarose gel. The primers used for PCR amplification of the endogenous ApoAII were Ex 1-1221 S and Rev Cla/Not, whereas for the pApo-wt Ex 1-1221 S and Not/Cla rev.

For the “in vitro” approach, we used the mouse α -tropomyosin region encompassing exon 2 and 3 separated by a synthetic intron sequence (Deirdre et al., 1995) (PY7 vector, a kind gift of Dr. C.W.J. Smith from Cambridge University) as model substrate where to test the role and the relevance of constitutive and accessory splicing elements of ApoAII exon 3 through “in vitro” splicing assays.

To this end, the wild type or mutated versions of the whole ApoAII exon 3 and its polypyrimidine tract were cloned in the PY7 construct replacing the corresponding elements of the α -tropomyosin gene. The effects of these changes were tested by “in vitro” splicing experiments under standard conditions.

Influence of the 3' splice site strength on splicing of ApoAII exon 3

In vivo effects of (GT)16 tract deletion and replacement

Previous experiments in our group have shown that when the T tract is removed from human CFTR intron 8, total exon 9 exclusion occurs (Niksic et al., 1999). Moreover, it was shown that an increase of the number of (GT) dinucleotide repeats reduces the efficiency of exon 9 inclusion. This splicing pattern is in contrast with what is observed in the ApoAII context where exon 3 inclusion reaches a proportion of 95%. Therefore, in the context of ApoAII the (GT) dinucleotide repeats seem to work as a functional polypyrimidine tract. In order to test this hypothesis we deleted the whole (GT) dinucleotide repeats from the ApoAII intron 2 [(Δ gt) in Fig.2.7]. As a control, a sequence containing (CA)16 dinucleotide repeats was used to replace the (GT)16 tract maintaining the same length (ca)x(gt), (Fig. 2.7). We found that both the deletion and the replacement of the (GT) tract resulted in total exon 3 exclusion.

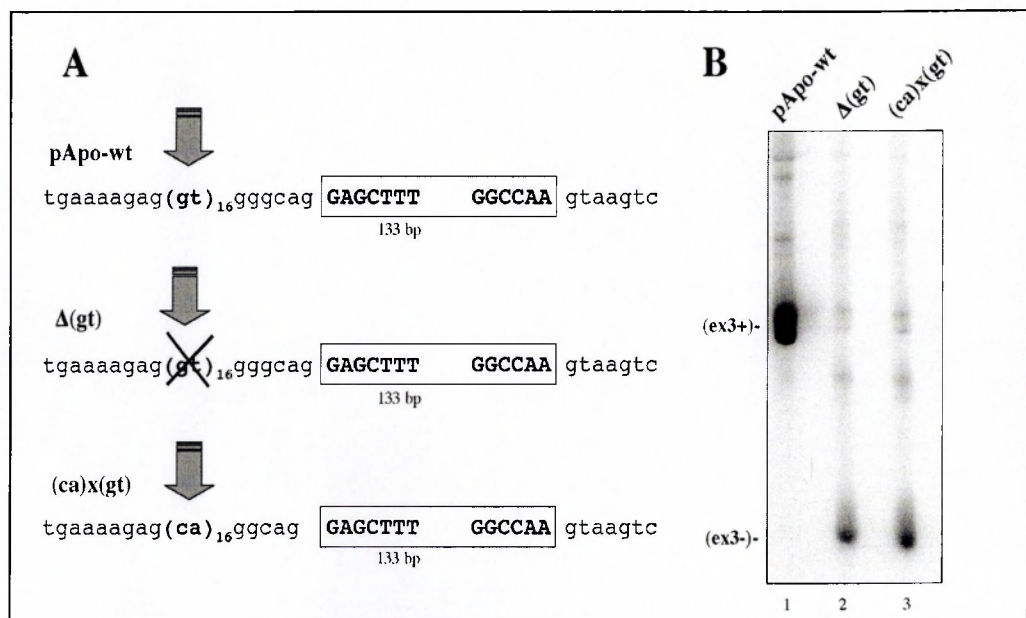


Figure 2.7. In vivo effect of GT deletion/replacement. A) Partial sequence of exon 3 and flanking introns showing the modifications within the 3' splice site. B) Analysis by denaturing polyacrilamide gel electrophoresis of pre-mRNA splicing pattern of human Apo AII constructs in Hep 3B cells. The relevance of the GT tract at the 3' splice site of exon 3 was assessed by deletion and replacement by CA tract.

On one hand, these results indicate the relevance of the (GU)₁₆ dinucleotide repeats as a functional “polypyrimidine tract” in the ApoAII context. On the other hand, previous studies on CFTR exon 9 and the results with EDA model not only suggest that the (GU)₁₆ tract alone is a weak polypyrimidine tract but also that the ApoAII exon 3 context should contain specific cis-acting elements that may counteract the weakness of its polypyrimidine tract.

A 45 kDa protein interacts with the 3' splice site of ApoAII exon 3

Buratti et al have reported the identification of TDP-43 as a splicing regulator that specifically binds the polymorphic (GU) repeats within the 3'-splice site of CFTR exon 9 (Buratti et al., 2001). Originally, TDP-43 was isolated following its binding to the pyrimidine rich motifs of Human Immunodeficiency Viruses type I TAR DNA sequence where is likely to be critical in the control of gene expression (Ou et al., 1995). In fact, it is a potent inhibitor of HIV-1 gene expression, during the early phase the HIV-1 cycle. The structural organization of this protein is similar to other RNA binding proteins: it contains three RRM (RNA recognition motifs) and a glycine rich domain. The group of cellular proteins to which TDP-43 belongs binds to a variety of different RNAs (Dreyfuss et al., 1993; Kenan et al., 1991). Interestingly, Ou et al. have demonstrated binding ability of this protein for the double stranded DNA, exclusively for polypyrimidine sequences containing at least 8 pyrimidines in the row (Ou et al., 1995). In contrast with these findings, previous studies in our group have shown that short RNA sequence consisting of alternated pyrimidine/purine, are successfully recognized and bound by TDP-43. Moreover TDP-43 can efficiently bind a number of (UG) repeats equal to or greater than six and there is a relationship between the number of (UG) repeats and the efficiency of binding (Buratti and Baralle, 2001).

In fact, UG but not U repeats within the intron 8 of human CFTR mRNA binds TDP-43.

Up to now, it is not clear which is the function, if any, of this cellular factor in the splicing regulation. It is possible that the increase in length of the UG tract would result in an increase of the binding affinity for this protein. Alternatively, the UG tract might bind more molecules of TDP-43, where number of the bound molecules could increase with the increasing number of UG repeats. In fact, the glycine rich domain of the TDP-43 suggests its ability to form protein-protein interaction.

Through its binding to UG sequence, TDP-43 might influence splicing assembling in two ways: its binding might indirectly mask poly U stretch modulating the binding efficiency of U2AF65 to the polypyrimidine tract or, alternatively, TDP-43 could directly be involved in the formation of new no favored protein-protein interaction with other proteins of the splicing machinery thus affecting exon 9 recognition. In any case, it is conceivable that binding of TDP-43 to the UG tract might lower exon 9 definition by acting as a “disturbing” element placed in proximity of its 3’ splice site junction.

TDP-43 has never been described as affecting the splicing process and its exact cellular function has remained elusive (Ou et al., 1995). Therefore, we wished to test if TDP-43 was also able to bind the (GU) repeats in the ApoAII context.

To this aim, UV-crosslinking experiments were carried out with a construct where the ApoAII exon 2, intron 2 and exon 3 were cloned in the pBluescript SK vector under the T7 promoter control [Fig.2.9A, mgApo-wt(GT)16]. A mutant version of this construct where the (GT) repeats within the polypyrimidine tract were deleted was also generated [Fig. 2.9A, mgApo- Δ (GT)]. The two RNAs were in vitro transcribed in presence of the α -³²P-UTP and incubated with HeLa nuclear extracts. The RNA-protein interactions were made permanent by UV crosslinking exposure, and unprotected RNAs were degraded by RNase treatment. After

separation of the ^{32}P -labeled proteins by SDS-PAGE, the gel was dried and exposed to autoradiographic films.

Comparing the pattern of proteins interacting with the two RNAs, it was apparent that among the numerous proteins that could be crosslinked to the labeled RNAs a cellular protein of apparent molecular weight over 47.5 kDa could be observed only in the mgApo-WT RNA which contained the (GU) repeats but not mgApo- Δ GT RNA that lacked the repeats (Fig.2.9B). This results were consistent with what was observed for the CFTR exon 9 GU tract (Buratti and Baralle, 2001; Buratti et al., 2001), in which crosslinked TDP-43 migrated near the 47.5 kDa marker. In order to confirm the identity of the 47.5 kDa protein, UV-crosslinking assays followed by immunoprecipitation with polyclonal antibodies anti-TDP-43 were carried out. Figure 2.9C shows that the anti-TDP-43 antibodies were able to immunoprecipitate a 47.5 kDa protein-mgApo-wt RNA complex, whereas no proteins were immunoprecipitated with mgApo- Δ GT RNA. Therefore, this experiment confirmed that TDP-43 is able to bind the (GU)₁₆ repeats in the ApoAII intron 2 context.

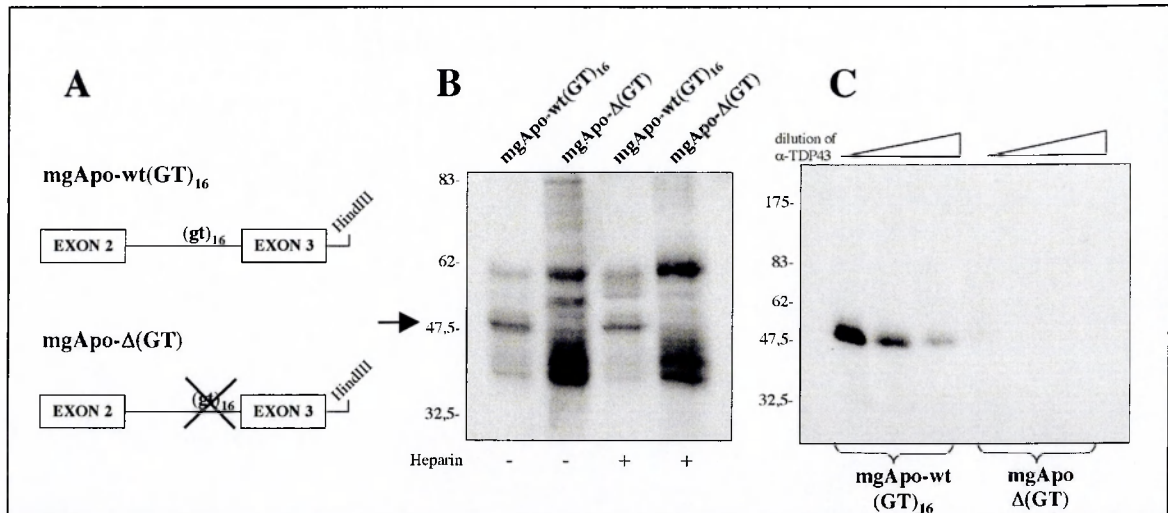


Figure 2.9. Effect of the GT tract deletion on RNA/protein interaction A) Scheme of two Apo AII minigenes spanning from exon 2 to intron 3 (up to base 2022) with or without the TG tract within intron 2. B) UV crosslinking reactions of RNA transcripts synthesized from the constructs by the T7 RNA polymerase and linearized with HindIII. Transcripts mgApo-wt(GT)₁₆ and mgApo-Δ(GT), were incubated with HeLa nuclear extract either under standard conditions or using Heparin as unspecific competitor concentration at the final concentration of 5 μg/μL. After UV irradiation, samples were RNase digested, resolved on SDS-PAGE and autoradiographed. Positions of molecular weight marker are indicated. C) Immunoprecipitation following UV cross-linking of labeled mgApo-wt(GT)₁₆ and mgApo-Δ(GT) transcripts with increasing dilutions of anti TDP-43. Positions of molecular weight marker are indicated.

5' splice site involvement in the recognition of the ApoAII exon 3

The analysis of the strength of the splice sites showed that the 5' splice site of the ApoAII exon 3 has a good match with the consensus. In order to test the relevance of the 5' splice site for ApoAII exon 3 definition, it was mutagenized away from the consensus (Fig. 2.10A, pApoAII $\Delta 5'$ construct). This experiment showed that the deletion of the wild-type 5' splice site resulted as expected in the exon 3 exclusion in the ApoAII mRNA and not in intron retention (Fig. 2.10B) suggesting the “exon definition” as the mechanism by which the exon is recognized and spliced (Robberson et al., 1990). Moreover, this result suggests that the efficient inclusion of exon 3 might be due to the presence of exonic regulatory elements similarly to what happens in other examples of splicing according to the “exon definition” model.

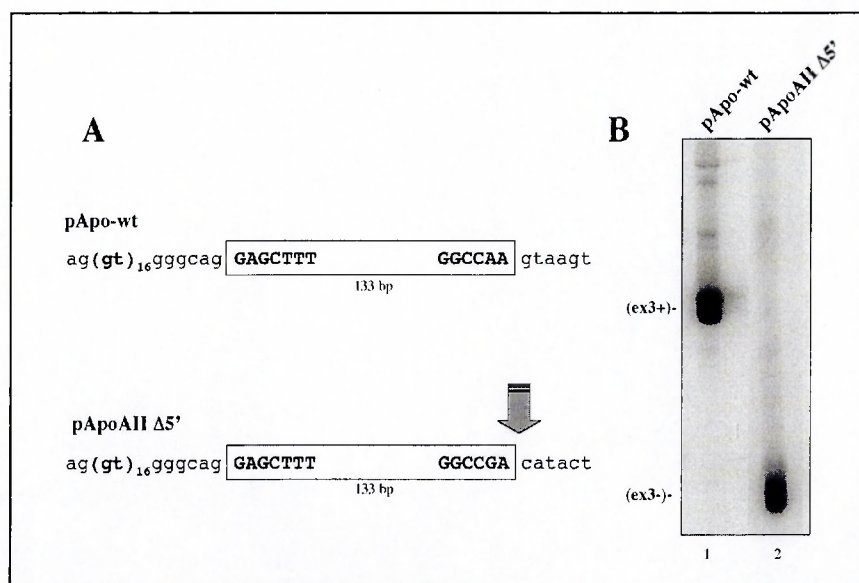


Figure 2.10. Effect of ApoAII exon 3 5' splice site mutation on Apo exon 3 splicing. A) Partial sequence of exon 3 and flanking introns showing the mutations within the 5' splice site. The 5' splice site of exon 3 was mutagenized as follows (CCAA/gtaagt \Rightarrow CCGA/cataact). B) Analysis by denaturing polyacrilamide gel electrophoresis of pre-mRNA splicing pattern of human Apo AII constructs in Hep 3B cells.

This radical change on the 5' splice site gives little information about more subtle modulating elements. Hence we tried to lower the strength of the human ApoAII exon 3 5' splice site (SSPNN score=1.0) by replacement of the splice junction and the 266bp of intron 3 with the 5'splice site of human CFTR exon 9 (SSPNN score 0.83) and the 266bp of its flanking intron 9 (Fig. 2.11A, construct pApo/hCF5'). In this construct, the previously characterized ISS regulatory element of CFTR intron 9 was not included (Pagani et al., 2000).

As a control, another construct was prepared where the ApoAII mouse 5' splice site (SSPNN score=1.0) was used for the replacement (Fig. 2.11A, construct pApo/mApo5'). Transient transfections of the pApo/hCF5' construct resulted in an increase up to 40% of ApoAII exon 3 skipping in comparison with pApo-WT (Fig. 2.11B). Interestingly, the replacement with the mouse ApoAII 5' splice site and its intronic flanking sequences resulted also in an increase up to 20% of exon 3 skipping (Fig. 2.11B).

On one hand, the experiments with pApo/hCF5' construct further suggest that the 3' splice site of human ApoAII exon 3 is weak. In fact, it is apparent that the authentic strong 5' splice site is necessary to balance a weak 3' splice site. On the other hand, the experiment with pApo/mApo5' construct suggests that, in comparison with the murine counterpart, other accessory regulatory elements might be present also within the human ApoAII intron 3 and that support exon 3 definition.

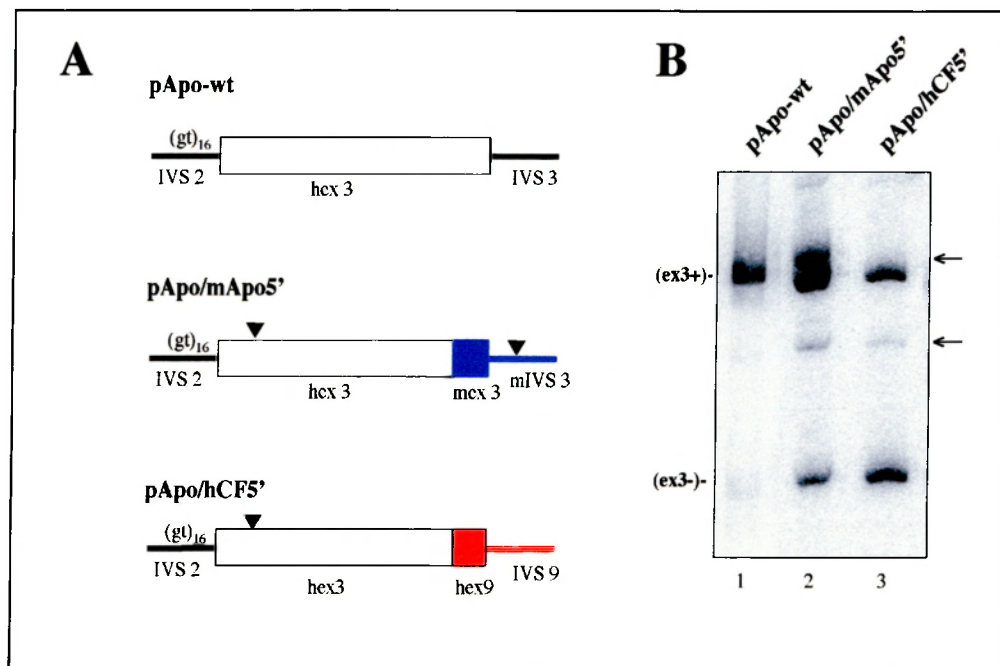


Figure 2.11. Effect of the human ApoAII 5' splice site replacement by mouse ApoAII 5' splice site and human CFTR 5' splice site. A) The main features of the constructs used for transient transfection in Hep 3B cells are shown. Mouse ApoAII 5' splice site and CFTR 5' splice site are shown in blue and red, respectively. B) Analysis by denaturing polyacrilamide gel electrophoresis of pre-mRNA splicing pattern of hybrid Apo AII constructs in Hep 3B cells. The horizontal arrow on the RT-PCR results indicates aberrant splicing products originated from cryptic 3' and 5' splice sites located in the hybrid exon and intron respectively (▼)

Mapping of regulatory elements within ApoAII exon 3

The different splicing pattern exhibited by human CFTR exon 9 and human ApoAII exon 3 cannot be explained simply with a different strength of the splice sites. In fact, the results described above suggest the presence of regulatory elements possibly localized both within exon 3 and in the downstream intron. Since the splicing regulatory elements of CFTR exon 9 were previously characterized, to identify which sequences in the human ApoAII exon 3 were important for its definition we generated hybrid constructs between the ApoAII exon 3 and CFTR exon 9. To this aim, the ApoAII exon 3 was progressively replaced with CFTR exon 9 and the constructs were transfected into Hep3B cells.

The first series of hybrids generated was a set of five CFTR/ApoAII constructs, in which the exon 3 was progressively replaced by CFTR exon 9 and its 3' splice site with the allelic

configuration (GU)11T5. The constructs were named by indicating the number of ApoAII exon 3 nucleotides replaced by CFTR exon 9 (i.e. 18, 39, 60, 87, 113) (Fig.2.12).

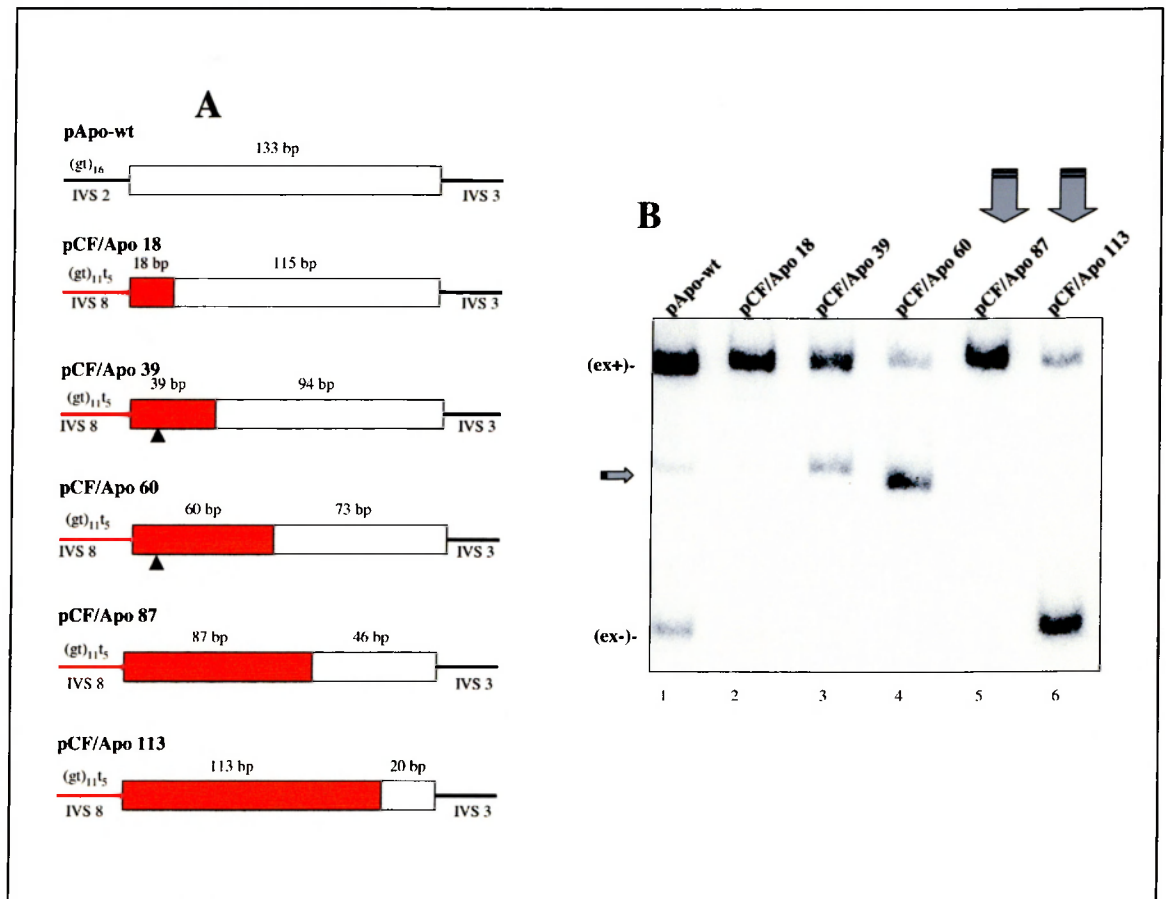


Figure 2.12. Effect of the progressive Apo AII exon 3 replacement by CFTR exon 9. A) The schemes of the CFTR IVS8-exon 9 (red)/ApoAII exon 3-IVS3 (white) hybrids constructs used for transient transfections in Hep 3B cells are shown. The size of such replacements is also indicated. B) Analysis of pre-mRNA splicing of CFTR/Apo AII hybrids constructs by RT-PCR. After the 30 cycles of PCR, radioactive CTP was added to each reaction and other 5 cycles more were added. The PCR products were separated on 7M urea 8% acrylamide gel. The horizontal arrow on the RT-PCR results indicates aberrant splicing products originated from cryptic 3' splice sites located in the hybrid exon 3 (▲).

In comparison with the splicing pattern of pApo-wt, pCF/Apo18 construct showed an increase of exon 3 inclusion up to 97% indicating, as expected, that the CFTR (GT)₁₁T₅ polypyrimidine tract is stronger than the ApoAII (GT)₁₆ polypyrimidine tract.

Then, pCF/Apo39 and pCF/Apo60 constructs showed the activation of cryptic 3' splice sites introduced with the extension of the CFTR sequence from nucleotides 39 to 60. This cryptic splice site has been previously characterized (Niksic et al., 1999). More interestingly, the two last constructs showed a striking difference in their splicing pattern. The transfection of pCF/Apo87 construct showed 100% inclusion of exon 3, whereas pCF/Apo113 construct showed 70% exclusion of exon 3 (Fig. 2.12B lanes 5 and 6). This result suggests the presence of a regulatory element localized within the 26bp in between pCF/Apo87 and pCF/Apo113 constructs. The different splicing pattern of pCF/Apo87 and pCF/Apo113 constructs might be due to the removal of an enhancer element placed within the 26bp of ApoAII exon 3.

Nevertheless, in principle, it should be noticed that the different splicing pattern might also depend on the addition of a silencer element introduced with the 26bp of the CFTR exon 9. However, considering that previous mapping studies did not highlight the presence of any exonic CFTR regulatory element in this 26nt-range, we focused the attention on the enhancer-like sequence possibly removed from ApoAII exon 3.

To identify sequences within this 26bp region that contributed to efficient exon 3 splicing, we constructed three 21bp-overlapping deletions encompassing the 26bp sequence (Fig.2.13A). The three constructs (Δ 1843-63, Δ 1854-74, Δ 1866-86) were transiently expressed in Hep3B and the RNA was analyzed by RT-PCR (Fig. 2.13B).

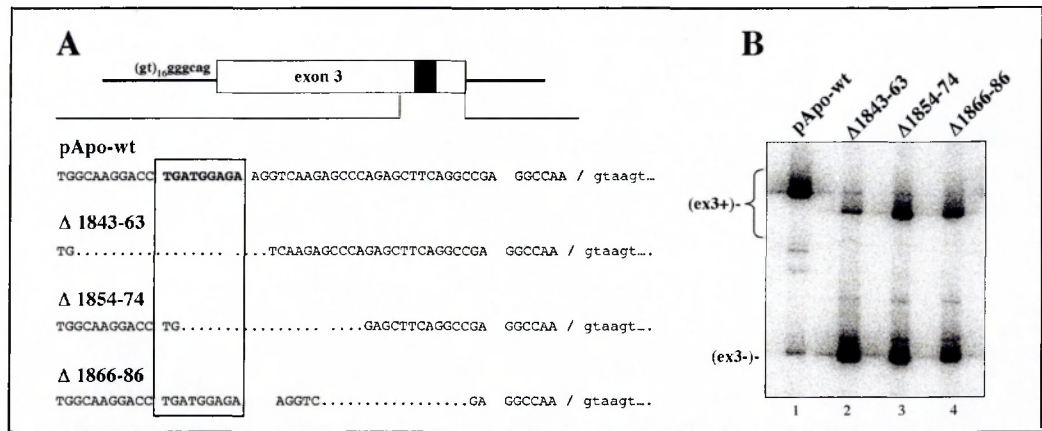


Figure 2.13. Effect of the overlapping deletions within Apo AII exon 3. A) Scheme and sequence of the overlapping 21 nt-deletions within exon 3. The rectangular box indicates the place in exon 3 where the putative ESE is located. Only Δ1843-63 construct deleted the whole ESE region. B) Analysis by denaturing polyacrylamide gel electrophoresis of pre-mRNA splicing of the constructs carrying the deletions by RT-PCR. After the 30 cycles of PCR, radioactive α -³²P-CTP was added to each reaction and other 5 cycles more were added. The PCR products were separated on 7M urea 8% acrylamide gel.

Whereas the Δ1854-74 and the Δ1866-86 deletions resulted in an increase of the exon 3 skipping up to 60% and 55%, respectively, the Δ1843-63 deletion caused a 90% exon 3 exclusion. Therefore, most of the ESE-like activity seems to be localized within the sequence interested by the first deletion. Hence, considering that a large number of ESEs are purine-rich, we outlined a 9bp-core region localized in the 1852-1860 nucleotide range (boxed sequence in figure 2.14C) that seems to belong to this class of enhancers.

At this point, in order to map finely which base(s) in such sequence was directly affecting exon 3 definition, point mutations within the 1852-1860 tract were carried out. Each base at a time was replaced by its complementary one generating nine different constructs that were used to transfect Hep3B cell line (Fig 2.14B).

While most of the substitutions did not have a significant effect on exon 3 processing in Hep3B cells, G92C, A99T and, more noticeably, A97T caused 15%, 60% and 85% of exon 3 skipping, respectively (Fig 2.14C).

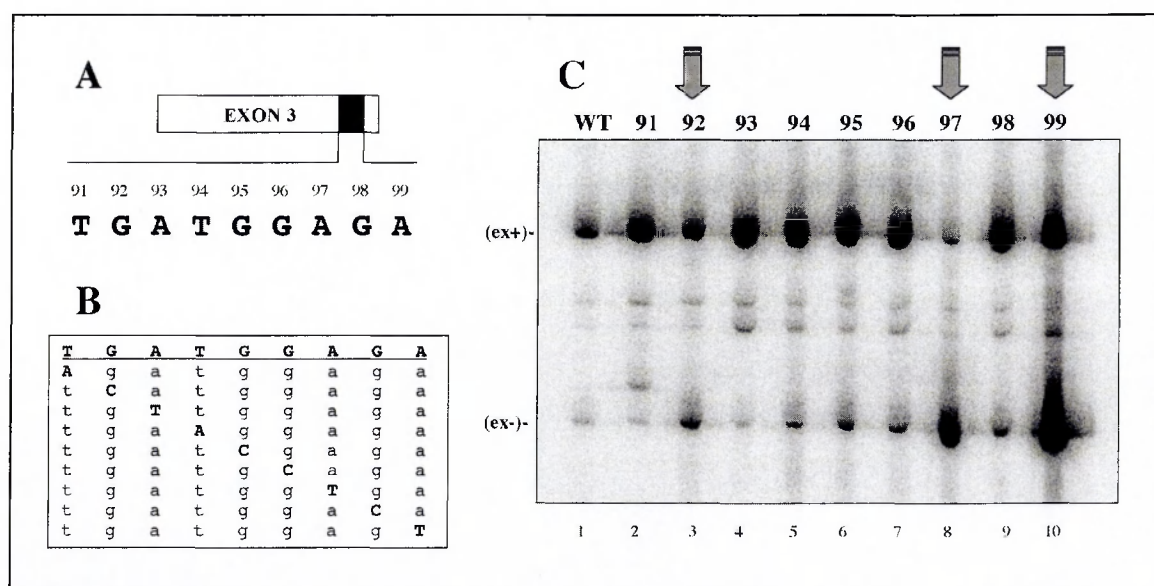


Figure 2.14. Effect of point mutations within the Apo AII exon 3 putative ESE. A) ESE sequence within the Apo AII exon 3. The number above each base corresponds to their position in the exon. B) scheme of the carried out point mutations. Briefly, each base at a time in the ESE was replaced by its complementary. Changes are indicated in bolded capital letters (diagonal). C) Analysis by denaturing polyacrilamide gel electrophoresis of pre-mRNA splicing of the constructs carrying the point mutations by RT-PCR. After the 30 cycles of PCR, radioactive CTP was added to each reaction and other and 5 cycles more were added. The PCR products were separated on 7M urea 8% acrylamide gel. Grey arrows indicate the strongest effect in exon3 skipping, shown by three of the constructs carrying point mutations at bases 92, 97 and 99.

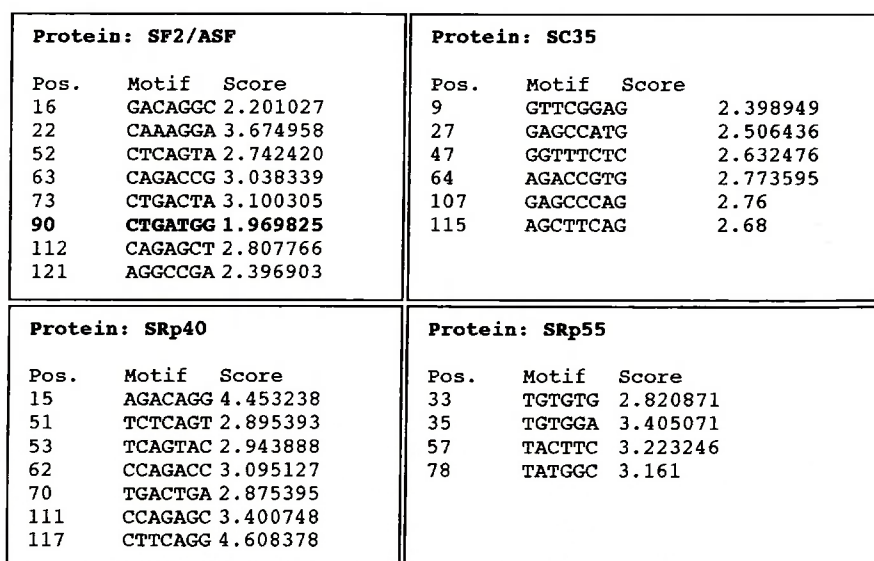
Screening of human ApoAII exon 3 for SR proteins consensus sequences

Often ESEs are *cis*-acting elements that work as binding sites for serine/arginine-rich (SR) proteins, a family of essential splicing factors that are also involved both in constitutive and alternative splicing. SR proteins promote splicing of adjacent introns. Although ESEs were initially grouped into two broad sequence-composition classes, purine-rich and AC-rich motifs, it is now apparent that under splicing conditions each SR protein recognizes specific, albeit degenerate and partially redundant, sequence motifs. Different studies aimed at obtaining a reliable prediction of genuine ESEs led to the calculation of score matrices derived from the functional consensus sequences of several SR proteins. In this way, it was possible to identify new ESEs and to show that high-score motifs are enriched in exons and are often

clustered in regions containing natural ESEs (Cartegni L, 2002; Cartegni et al., 2003; Fairbrother et al., 2002).

However, the predictive value of available score matrices was tested *in vivo* in different splicing models obtaining controversial results (Pagani et al., 2003). One explanation for the different exon 3 splicing pattern observed with A97T substitution might be that the nucleotide change might have differentially affected the SR protein profiles in the wild type sequence. Therefore, we calculated the available scoring matrices for SF2/ASF, SC35, SRp40, and SRp55 for the human ApoAII exon 3 wild type and A97T sequence using the scoring matrices that are available at the website <http://exon.cshl.org/ESE>.

In general, the results reported in Figure 2.15 show that the potential SR protein binding sites are clustered at the edges and in the center of exon 3. Then, in particular, one SF2/ASF potential site starting at nucleotide 90 (CTGATGG) was identified in the sequence range encompassing the 91-97 nucleotide core that we have found to work as ESE “*in vivo*”. On the other hand, no differences in the calculated matrix scores for the four SR proteins were apparent for A97T substitution in comparison with the wild type sequence. We also tested if the other substitutions we found to affect the splicing efficiency of ApoAII exon 3 (G92C and A99T) might be in correlation with the alteration of some matrix score. However, also in this case we were not able to highlight any difference in the calculated matrix scores (data not shown).



75

As already mentioned the predictive value of available score matrices was not absolute at least when tested “in vivo” in different splicing models (Pagani et al., 2003; Sironi et al., 2002). Therefore the observations made for the ApoAII exon 3 did not necessarily exclude splicing factor interactions in this region and testing these sequences experimentally was necessary.

Interaction of SR proteins with ApoAII 9nt-ESE

In order to identify the trans-acting factors able to bind the 9nt enhancer like sequence within ApoAII exon 3, both EMSA and UV-cross linking assays were performed. For the EMSA experiment, two in vitro transcribed RNAs containing either the 9nt wild type core region of the ApoAII exon 3 (ESE-wt) or not (Δ ESE) were incubated with HeLa nuclear extract in presence of the non-specific competitor heparin (see materials and methods) (Fig.2.16A). As shown in Fig. 2.16B, ESE-wt RNA showed a broad band of shifted material, whereas no significant shift was observed with the Δ ESE RNA. Therefore, this result suggests that one or more proteins might interact specifically with the wild type ApoAII exon 3 across the 9nt-sequence.

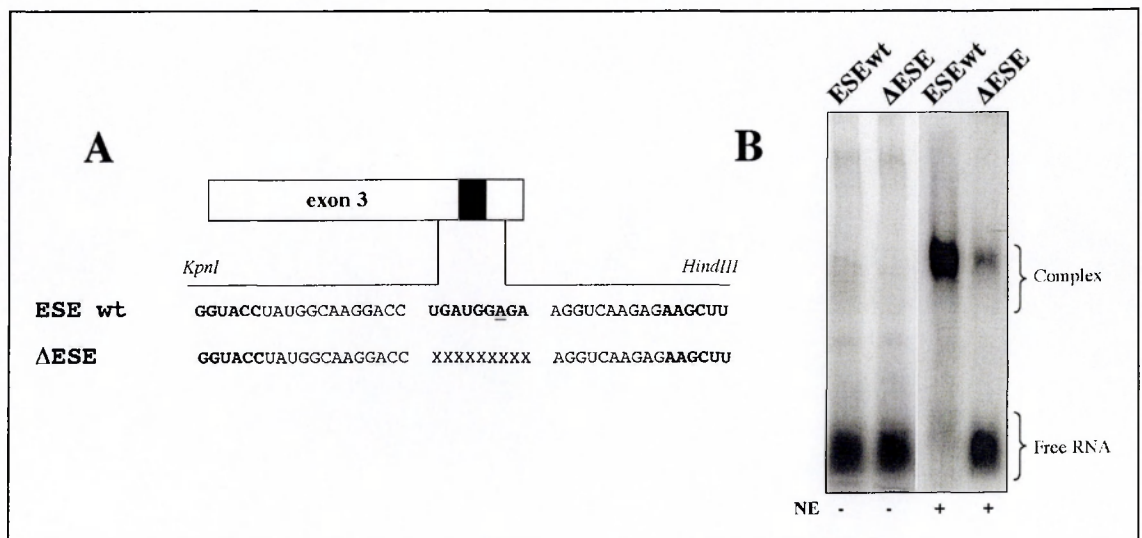


Figure 2.16. Electro mobility shift assay of ApoAII exon 3 ESE. A) Scheme of the constructs carrying wild type (ESE wt) or mutated (Δ ESE) exonic splicing enhancer of ApoAII exon 3 used for the EMSA. B) Radiolabeled RNAs were incubated with HeLa nuclear extract (NE) for 20 min at RT. Complexes were then fractionated on a 4% non-denaturing polyacrylamide gel. The position of bound or free RNA is indicated.

The nature of the protein(s) that bind to the wild type ApoAII exon 3 across the 9 nt-sequence was then investigated by UV cross-linking of RNA-protein complexes (see materials and methods).

³²P-labeled ESEwt, ΔESE, A97T RNA probes (Fig. 2.17A) were incubated with HeLa nuclear extract and were cross-linked to proteins by exposure to UV light; the resulting ³²P-labeled protein(s) were separated by SDS-PAGE. Fig.2.17B shows no differences in the pattern of UV cross-linked proteins obtained with these three different constructs.

The fact that no differences in the UV crosslinking have been observed might be ascribed to the high background of proteins in the molecular weight range in which the SR proteins migrate, also considering the fact that they are not abundant proteins.

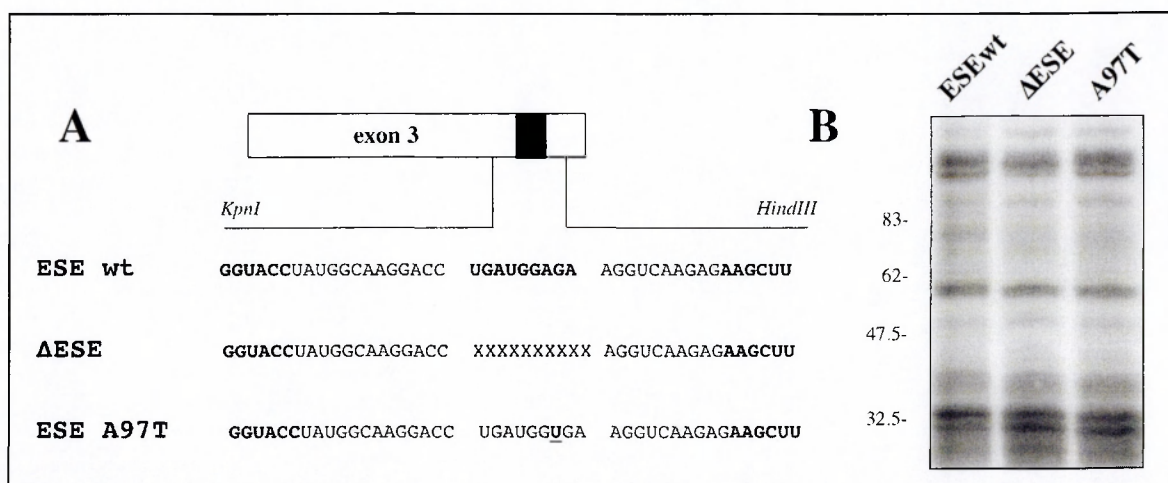


Figure 2.17. Effect of ESE deletion or mutation on RNA/protein interaction A) Schematic representation of the plasmids used, wild type (ESE wt), deletion (ΔESE) and point mutated (ESE A97T) exon 3 enhancer. B) 10% PAGE-SDS analysis of UV cross-linking assay using HeLa nuclear extract with uniformly labeled RNA of all three constructs linearized with the restriction enzyme *HindIII*.

As we have not observed differences in the protein pattern by UV-crosslinking, in order to test the possible recruitment of different SR proteins by ApoAII 9nt-ESE, UV-crosslinking assays followed by immunoprecipitation with monoclonal antibodies were carried out. The constructs used for these experiments were ESE-wt, ESE A97T and another mutant of ESE, Δ 1843-63, with a deletion of 21nt including the 9nt-ESE (Fig. 2.18A).

To outline possible differences in the ability to interact with specific members of the SR proteins family by ESE-wt RNA in comparison with ESE A97T and Δ 1843-63 RNAs, the UV-crosslinked samples were subsequently immunoprecipitated with the mAb 96 anti-SF2/ASF monoclonal antibody (Fig. 2.18B) and anti-SC35 monoclonal antibody (Fig.2.18C). Moreover, an ESE derived from the human fibronectin EDA was in vitro transcribed to be used as a positive control for ASF/SF2 and SC35 interactions (Buratti et al, Mol Cell Biol in press).

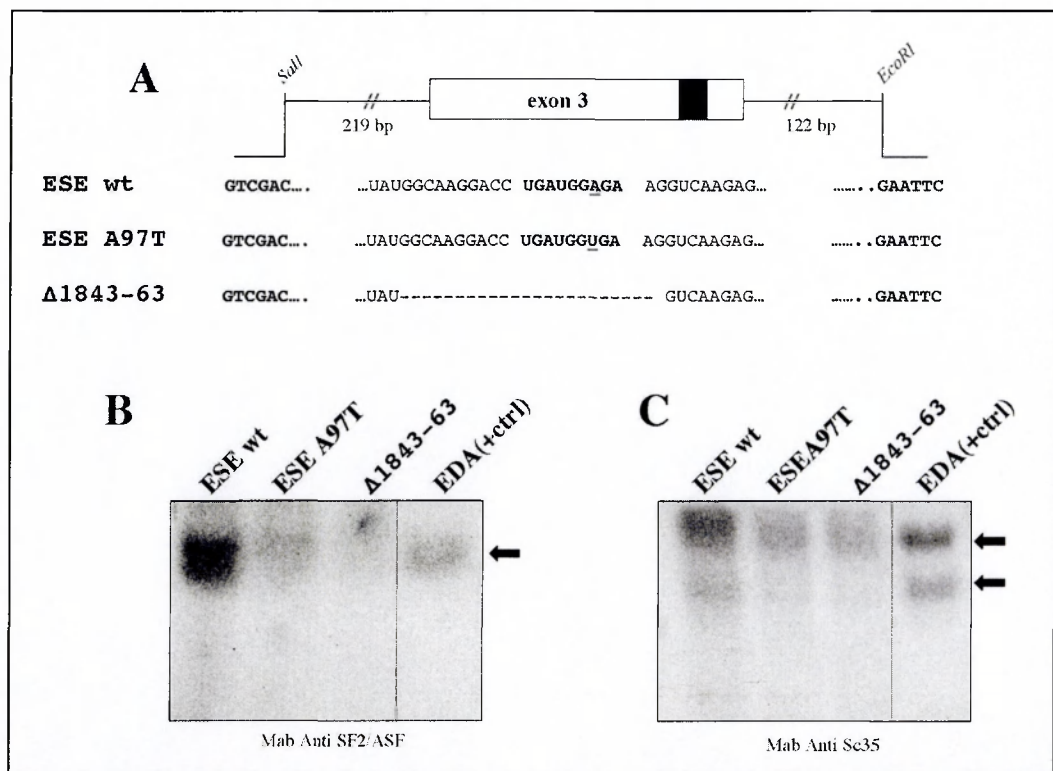


Figure 2.18. Specific interaction of SR proteins with the ESE within Apo AII exon 3. A) Partial sequence of the constructs used for UV crosslinking followed by immunoprecipitation. Three constructs containing exon 3 (with the intact, point mutated and deleted ESE) plus portions of the flanking introns 2 and 3 are shown. B) 10% PAGE-SDS analysis of immunoprecipitation with an anti-SF/ASF2 monoclonal antibody following UV cross-linking of labeled RNAs: ESEwt, ESEA97T and Δ1843-63 along with EDA as a positive control. The arrow indicates the 36-38 kDa immunoprecipitated product at the same level of the positive control. C) 10% PAGE-SDS analysis of immunoprecipitation with an anti-SC35 monoclonal antibody following UV cross-linking of labeled RNAs: ESEwt, ESEA97T and Δ1843-63 along with EDA as a positive control. The arrows indicate a 35 and a 40 kDa immunoprecipitated product at the same level of the positive control.

The results show a protein that can be specifically immunoprecipitated by the mAb 96 (Fig.2.18B, lane 1) and by anti-SC35 (Fig.2.18C, lane 1) following UV crosslinking with nuclear extract of a labeled ESE-wt RNA, but not with the ESE A97T and Δ1843-63 control RNAs (Fig.2.18B and 2.18C, lane 2 and 3). Importantly, the proteins UV crosslinked with the migration control EDA-exon RNA and immunoprecipitated by anti-SC35 and mAb 96 have the same electrophoretic mobility of those present in ESE-wt. In this way, the identity of the proteins was confirmed. Thus, these results provide an evidence that at least two SR proteins are able to interact with the sequence across the ESE sequence and it is also consistent with

the result obtained in the EMSA study where a clear difference in the RNA-protein complex formation with a construct containing the ESE wt and Δ ESE is observed.

Effect of SR proteins and hnRNP A1 overexpression on splicing of ApoAII exon 3

Both constitutive and alternative splicing of many genes are affected by the intracellular concentrations of antagonistic splicing factors of the SR family and hnRNPA1. Initially, to evidence the possible role of hnRNPA1 on ApoAII exon 3 splicing, we tested the effect of the overexpression of this protein by cotransfecting a plasmid carrying the hnRNPA1 coding sequence along with the pApo-wt construct. This construct was used because its transfection results in more than 90% of exon 3 inclusion in pre-mRNA.

As a control, two different SR proteins (ASF/SF2 and SC35), known to have opposing effects on splicing in comparison with hnRNP A1, were also cotransfected with pApo-wt construct. Figure 2.19B shows that, neither ASF/SF2 nor SC35 modified the ApoAII exon 3 splicing pattern, whereas hnRNPA1 caused a 20% of exon 3 skipping as compared with the empty expression vector. This result is consistent with other splicing models where hnRNPA1-overexpression shows to induce exon skipping (Caceres et al., 1994; Pagani et al., 2000).

It is more difficult to highlight the putative positive effect of SR proteins because the wild type construct shows already a very high exon 3 inclusion rate. We have used the Δ 1866-86 construct (scheme in Fig. 2.13A) because this deletion causes about 55% of exon 3 skipping without covering the 9 bp-core region of the ApoAII exon 3 ESE (Fig.2.19A). In consequence a positive effect of SF2 or SC35 interacting in the ESE region will be more visible. We have tested cotransfections of the Δ 1866-86 construct with constructs expressing the SR proteins SRp20, SRp55, SRp75, ASF/SF2 and SC35. Figure 2.19C shows that cotransfections of Δ 1866-86 with SRp55, SRp75, ASF/SF2 and SC35 promoted exon 3 inclusion when compared with the cotransfection with the empty vector. The effect is not dramatic but this

may be due to the fact that the 1866-86 deletion, occurring just 5nt downstream of the ApoAII exon 3 ESE core region, partially alters the RNA secondary structure of the region so interfering with the correct display of the ESE. The results obtained with SRp55 and SRp75 overexpression suggest the possibility that also these splicing factors might be involved in the mechanism underlying exon 3 splicing. Both ASF/SF2 and SC35, whose specific interaction to the ESE had already been demonstrated by UV crosslinking/immunoprecipitation (Fig.2.18 B and C), have a positive effect on exon 3 inclusion. On the other hand, SRp20 overexpression caused an increase of exon 3 exclusion (Fig.2.19B, 85%). Such an effect requires further experiments in order to map the binding site of this SR protein within the exon 3 and its flanking intronic sequences.

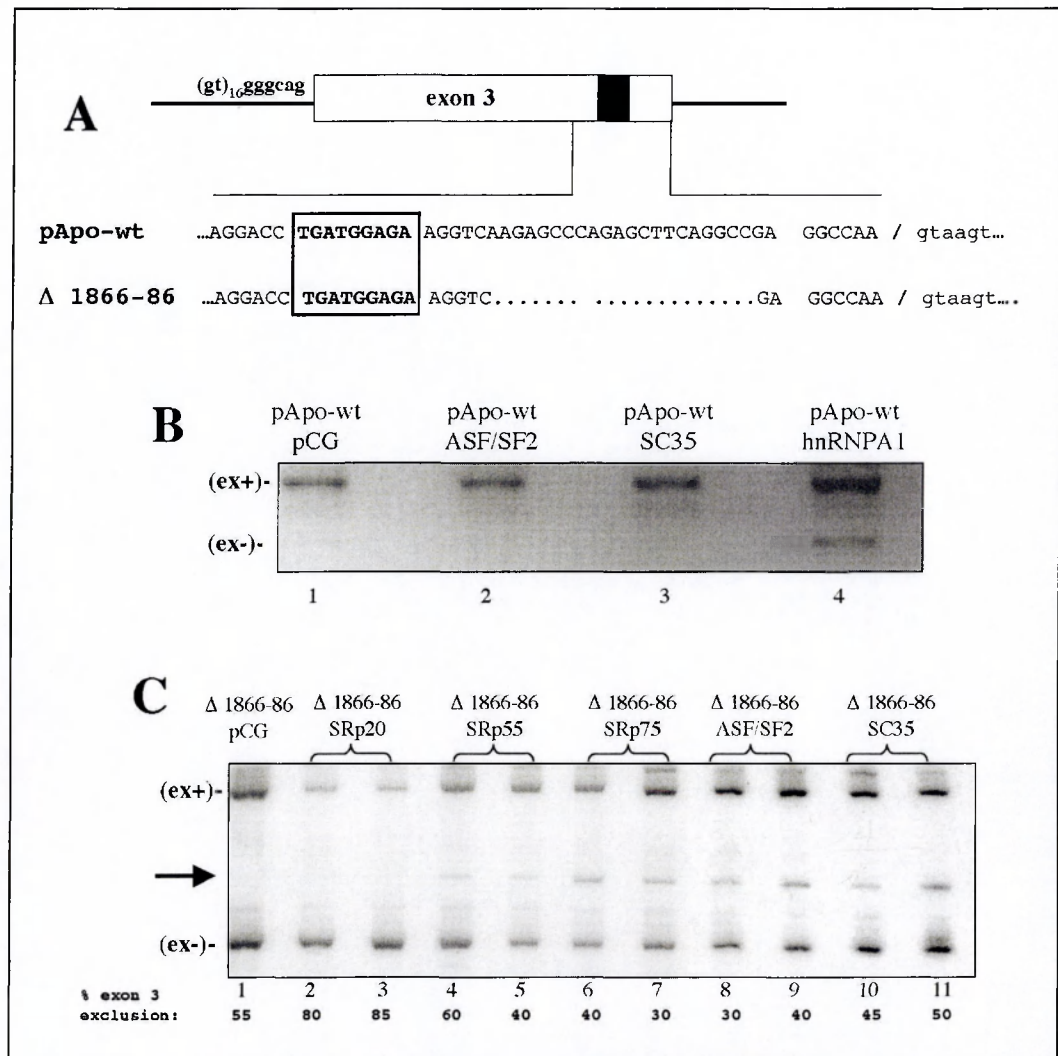


Figure 2.19. Effect of overexpression of some SR proteins and hnRNPA1 on Apo AII exon 3 splicing A) Partial sequence of the constructs used for cotransfection with SR proteins or hnRNPA1. Both constructs present the intact ESE (box), but $\Delta 1866-86$ presents a 21-nt deletion that cause a 50% of exon 3 exclusion.. B) Splicing pattern analysis by agarose gel electrophoresis of the cotransfection of construct pApo-wt with 2 μ g of SR proteins (ASF/SF2, SC35) and hnRNPA1 coding plasmid. Proportion of exon 3 exclusion was determined by densitometry using the total densitometric units of 3+ and 3- amplicons as 100%. C) Splicing pattern analysis by denaturing polyacrilamide gel electrophoresis of the cotransfection of construct $\Delta 1866-86$ with 1 and 2 μ g of SR proteins 20, 55, 75, ASF/SF2 and SC35. These results represent three independent experiments. The arrow on the RT-PCR, C, indicates an aberrant splicing product originated from a cryptic 3'splice site located in the exon 3.

The position of ApoAII ESE within an internal exon is crucial for its function

Several studies have shown that splicing of pre-mRNAs containing weak 3' splice sites requires splicing enhancer elements for activity (Lynch and Maniatis, 1996; Ryner and Baker, 1991; Tian and Maniatis, 1992). Therefore, in order to verify the possible relationships between the 9nt-enhancer and the 3' splice site of ApoAII exon 3, we created a construct spanning from the (GT)₁₆ polypyrimidine tract to the end of the ApoAII exon 3 that replaced the corresponding region of the tropomyosin exon 3 (Fig. 2.20A, ESE-wt). Then, we generated a similar construct carrying a mutated version of the 9nt-ESE (Fig. 2.20A, ESE-A97T). Both constructs were used for *in vitro* splicing assays along with the original PY7 plasmid. If there was a connection between the ESE function and the 3' splice site of ApoAII exon 3, the splicing of ApoAII exon 3 should occur. Otherwise, no processing of the intervening sequence should be observed leading to the accumulation of RNA precursor. In contrast to what was observed “*in vivo*”, i.e. 90% of exon 3 skipping in presence of the A97T substitution, both ESE-wt and ESE-A97T were spliced apparently with the similar efficiency (Fig. 2.20B).

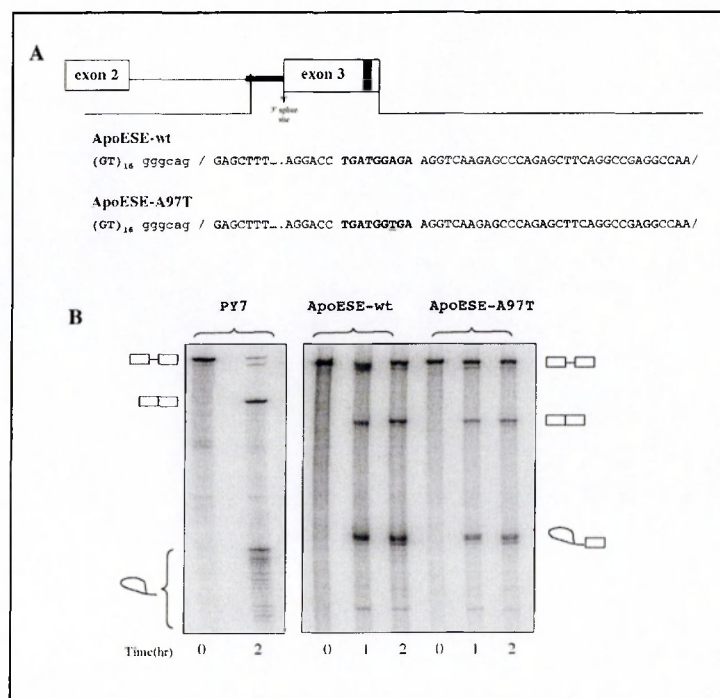


Figure 2.20. *In vitro* study of the effect of A97T mutation on the Apo AII exon 3 ESE A) Scheme of the two constructs used for *in vitro* splicing. Apo AII exon 3 (ESE-wt and ESE-A97T) and its polypyrimidine tract replaced the corresponding region of the tropomyosin exon 3 in the PY7 plasmid. B) *In vitro* splicing analysis by denaturing gel electrophoresis of the ESE A97T mutation effect. Labeled RNA was incubated with HeLa nuclear extract and the products of the splicing reaction were resolved on 7M urea 8% acrylamide gel. The initial transcripts and splicing intermediates are depicted on the sides of the gel and the incubation times are indicated at the bottom.

The observation that there were no differences in the splicing efficiency of both ApoESE-wt and ApoESE-A97T constructs *in vitro* suggested that the context of the PY7 construct might alter the ApoAII-ESE-dependant splicing of exon 3. In particular, the fact that the ApoAII ESE occurs within the “last” exon of the PY7 minigene system might affect the function of this ESE, normally placed within an internal exon.

In order to confirm this hypothesis (i.e., that the position of the ApoAII ESE within an internal exon is crucial for its function), we created a two-exon ApoAII system for “in vivo” experiments. For this purpose, two constructs were generated where intron 2 (pcDNApo-IVS2) or intron 3 (pcDNApo-IVS3) were reinserted in their natural position within the ApoAII cDNA. In this way, in the pcDNApo-IVS2 construct, intron 2 is in between of two exons where exon 1 and exon 2 forms the virtual “first exon” whilst exon 3 and exon 4 forms the virtual second and “last exon” (Fig. 2.21A).

On the other hand, in the pcDNApo-IVS3 construct, intron 3 is in between of two exons where exon 1, exon 2 and exon 3 form the virtual “first exon” whilst exon 4 alone forms the virtual second and “last exon” (Fig. 2.21A).

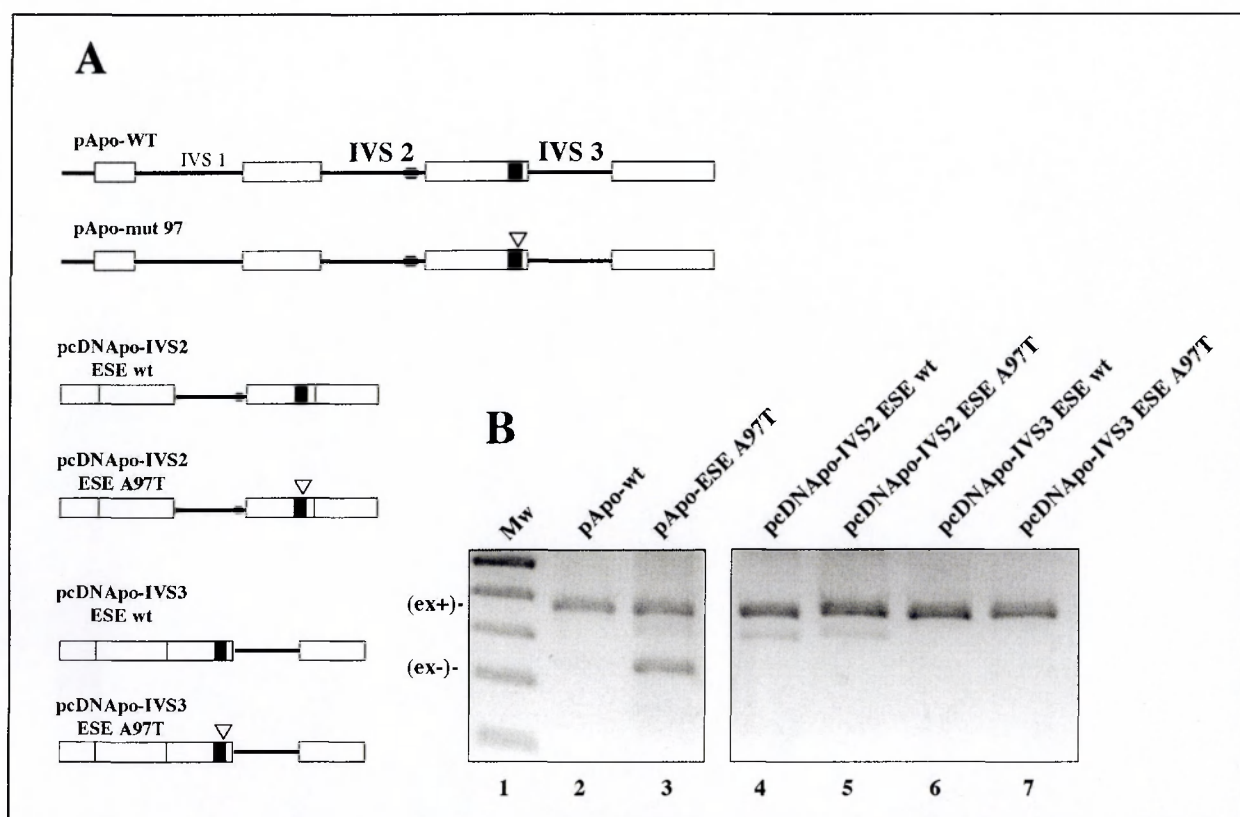


Figure 2.21. Relevance of the Apo AII exon 3 ESE position in *in vivo* studies. A) Scheme of the constructs used for transient transfection in Hep 3B cells. Intron 2 was introduced in its natural position in the Apo AII cDNA carrying the wt or the mutated A97T ESE (▽) pcDNApo-IVS2 ESEwt and pcDNApo-IVS2 ESE A97T respectively. Intron 3 was introduced in its natural position in the Apo AII cDNA carrying the wt or the mutated A97T ESE, pcDNApo-IVS3 ESEwt and pcDNApo-IVS3 ESE A97T respectively. B) Splicing pattern analysis by RT-PCR on 2% agarose gel. Full intron processing was obtained with all the constructs generated.

Afterwards, using pcDNApo-IVS2 and pcDNApo-IVS3 as templates, two mutants were created carrying the A97T mutation that was previously shown to inactivate the ApoAII exon 3 ESE (Fig. 2.21A). Therefore, in the pcDNApo-IVS2-A97T construct, the mutated ESE was localized within the virtual second and “last exon”, whereas in the pcDNApo-IVS3-A97T construct, the mutated ESE was localized within the virtual ‘first exon’ (Fig. 2.21A).

In the natural context (i.e. within an internal exon), the ESE works properly and exon 3 is constitutively included in the pre-mRNA (Fig. 2.21 lane 2). When it is disrupted, exon 3 undergoes alternative splicing (fig 2.21 lane 3)

Similarly to what observed with the “*in vitro*” experiments, if the ESE function was independent from its localization, we would expect to find no processing of intron 2 or intron

3 “in vivo” when the ESE is mutagenized (A97T). Instead, shifting the context of the ESE to the first or the last exon, we have observed that both introns 2 and 3 are efficiently spliced out in spite of the disruption of the ESE (Fig.2.21 lanes 5 and 7). Overall, these results support the hypothesis that ESE-dependant mechanism of splicing works only if the ESE is placed in an internal exon.

As a final demonstration of this hypothesis, we transferred ApoAII exon 3 and its flanking intronic sequences into an heterologous system. We used the α -globin/fibronectin reporter system (PTB), a minigene system used in our laboratory to study alternative splicing of several exons. The wild type and A97T ApoAII exon 3 with its flanking intronic regions (174 nt of intron 2 and 116 nt of intron 3) were cloned into the third intron of the PTB construct (Fig. 2.22A).

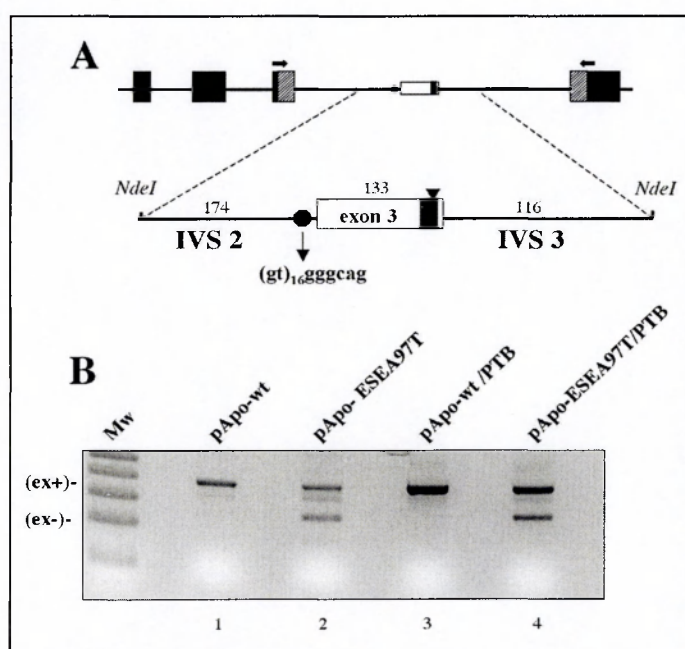


Figure 2.22. Apo AII exon3 ESE works only when located within internal exons. A) Apo AII exon 3 (with the ESE-wt or ESE A97T) and portions of its flanking introns was introduced in α -globin/fibronectin reporter system (PTB) as heterologous context. α -Globin, fibronectin EDB, and human Apo AII exons are indicated in black, grey, and white boxes, respectively. The black circle indicates the polymorphic GT tract. The primers used in the RT-PCR assay are indicated by the superimposed arrows. *NdeI* sites are indicated. The length of the relevant Apo AII fragments (intron 2, exon 3, and intron 3) is indicated. B) Splicing pattern analysis by RT-PCR on 2% agarose gel. The effect of the A97T mutation was recovered when the ESE is located in an internal exon.

Transfections of pApo-wt/PTB construct resulted in full inclusion of exon 3 in mRNA while pApo-ESEA97T/PTB construct showed exon 3 skipping (Fig. 2.22B lanes 3 and 4).

These results show that the ESE works also when ApoAII exon 3 is placed in between of two other exons in a heterologous system, so confirming that its effects are evident only when it is located within an internal exon.

Moreover, these results might indicate the presence of other splicing regulatory elements within the flanking intronic regions of ApoAII exon 3.

Order of intron removal in the ApoAII gene

The peculiar distribution of different cis-acting elements within the constitutive ApoAII exon 3 and its flanking introns led us to consider that it might reflect a splicing mechanism involving a particular order of intron removal. We determined the order of intron splicing in the ApoAII gene using an RT-PCR-based approach described by Kessler et al. (Kessler et al., 1993). In this experiment, only the observation of the partially spliced intermediates is instructive, since products containing all the introns may derive from the primary RNA transcript and are uninformative for the order of intron splicing. Figure 2.23A shows the strategy used to understand the order of intron removal. We used two sets of primers: in the first set, the forward primer was located in ApoAII exon 1 and the reverse primer in intron 3, whereas the second pair included a forward primer in intron 2 and a reverse primer in exon 4. With the first set of primer, a 610bp PCR products was observed along with the no spliced pre-mRNA spanning from exon 1 to intron 3 (771 bp). The 610 bp band was found to correspond to exon 1-exon 2-intron 2-exon 3- intron 3 pre-mRNA (Fig 2.23A right panel). These results seems to indicate that in the ApoAII gene, intron 1 is preferentially spliced before intron 2. In contrast, the second set of primers amplified the spliced intron 2-exon 3-exon 4 intermediate (478 bp), with traces of the unspliced product (875 bp) (Fig 2.23B right panel). These results seems to indicate that in the ApoAII gene, intron 3 is preferentially spliced before intron 2. Therefore, taken together these data suggest that the intron 2 is the last intron in being processed, whereas the order removal of intron 1 and 3 is not clear and will be tested in future experiments with different combinations of primers.

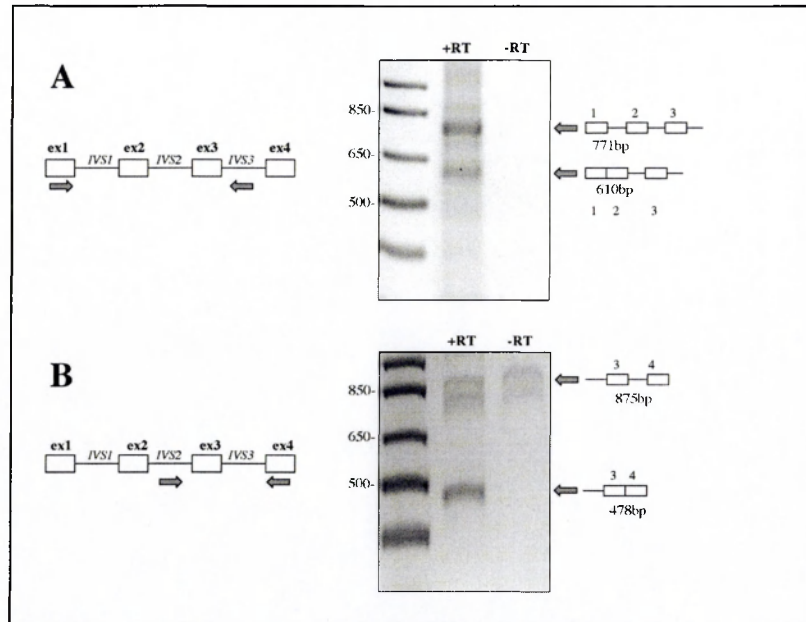


Figure 2.23. Splicing order of ApoAII introns 1, 2 and 3. A) RT-PCR products obtained for the first set of primers, separated on a 2% agarose gel. For each reaction, a PCR amplification was performed without prior reverse transcription (-RT) to exclude DNA contamination. The bands correspond to the unspliced product (771 bp) and the spliced exon 1-exon 2-intron 2-exon 3-intron 3 (610 bp), demonstrating splicing of intron 1 before intron 2. B) RT-PCR products obtained for the second set of primers separated on a 2% agarose gel. The bands correspond to the unspliced product (875 bp) and the spliced intron 2-exon 3-exon 4 (478 bp) demonstrating splicing of intron 3 before intron 2.

Mapping of regulatory elements within intronic regions flanking ApoAII exon 3

The possibility of the existence of splicing regulatory elements within the intronic regions flanking ApoAII exon 3 is suggested both by the experiments with human/mouse hybrid and by all the data related to the effect of the ESE within exon 3 in relationships with the simultaneous presence of upstream and downstream introns. In order to verify the presence of intronic regulatory elements and map their position, a series of constructs was prepared.

Initially, we screened the exon 3 flanking intronic regions looking for possible cis-acting elements whose sequence might resemble previously described modulators of splicing. We noticed the presence of three G runs in intron 3 just 10 bp downstream of the exon 3 donor splice site (G1, G2 and G3 in Fig.2.24A). Previous studies have shown that G runs are distributed throughout human introns and represent an enhancer-like element which seems to

strengthen the 5' splice site of constitutive introns in chicken tropomyosin, human alpha-globin and human growth hormone. We have, therefore, explored the possible effects of G runs on the regulation of the splicing of the ApoAII exon 3. To this aim, we first created mutants where the first two and all the three G runs were disrupted by point mutations (legend of Fig.2.24B). However, the transient transfection of this mutant did not show significant changes of the ApoAII exon 3 splicing pattern in comparison with the pApo-WT construct. A slight effect was only observed when the last G run (i.e. G3) was mutagenized causing a nearly 10% of exon 3 skipping (Fig.2.24C). Thus, it is reasonable to exclude that these G runs within intron 3 have a significant role in the splicing of exon 3.

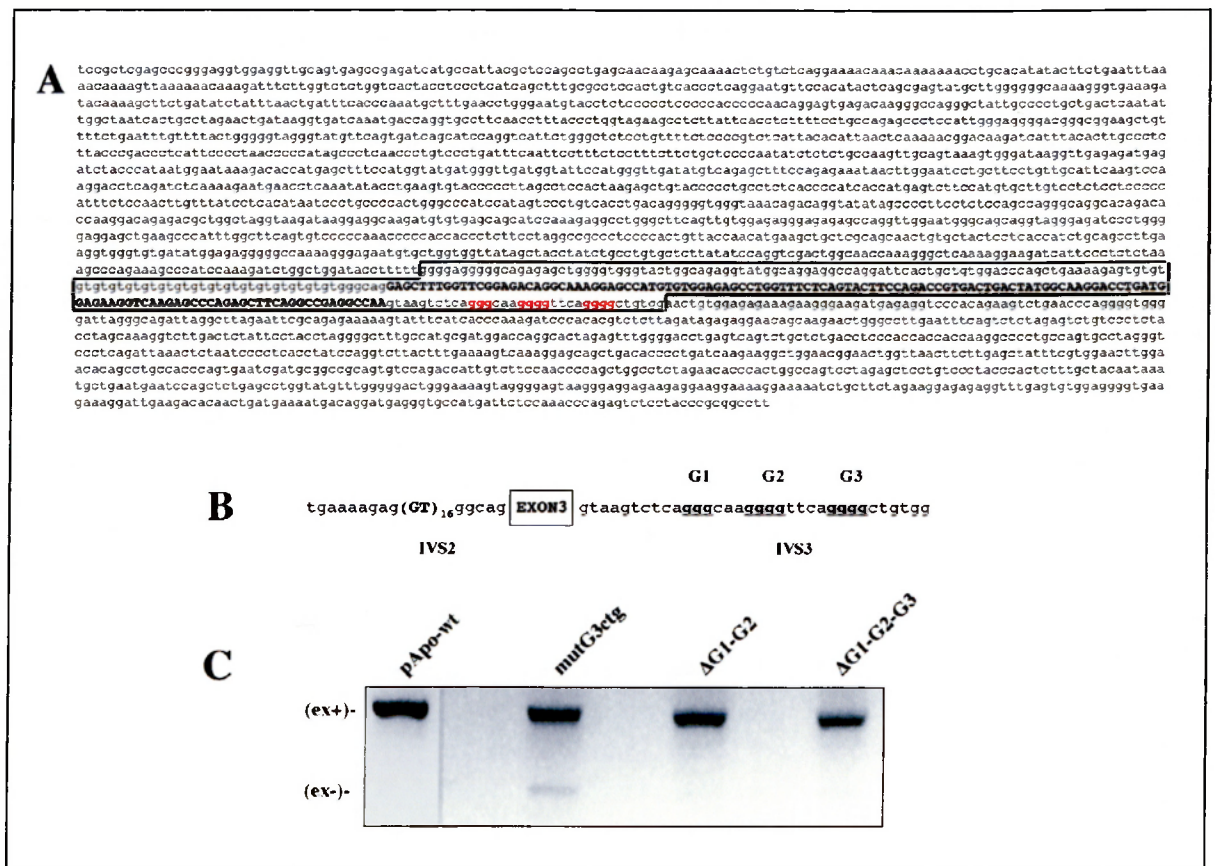


Figure 2.24. Effect of G runs mutations within Apo AII exon 3 flanking intron 3. A) ApoAII full sequence. In red are indicated runs of G localized in intron 3. The region encompassing the ApoAII exon 3 is bolded. B) Scheme of exon 3 plus flanking introns showing the G runs that were mutagenized, mutG3CTG: ggggctg->gagctta, ΔG1-G2: gggcaagggg->tcacaaagcgc, ΔG1-G2-G3: gggcaagggggttcaggggctg->tgtcaagcaltcatgcctgt C) RT-PCR analysis by agarose gel electrophoresis of the transfected constructs carrying mutations in the G runs within intron 3. No effect on exon 3 splicing is observed after mutagenesis of the G runs in the exon 3.

A series of deletions was then made within both introns in the pApo-wt and the pApo-A97T constructs first deleting 127nt-deletion of intron 2 ($\Delta 1554-1681$) along with a 91nt-deletion of intron 3 ($\Delta 1931-2022$) (Fig. 2.25A).

Figure 2.25B shows no differences in the exon 3 splicing pattern of the pApo-wt and the pApo-wt (2X Δ) constructs. However comparing the splicing pattern of pApo-A97T with pApo-A97T (2X Δ), we can observe that the simultaneous intronic deletions were able to counteract the effects of the ESE mutation within exon 3.

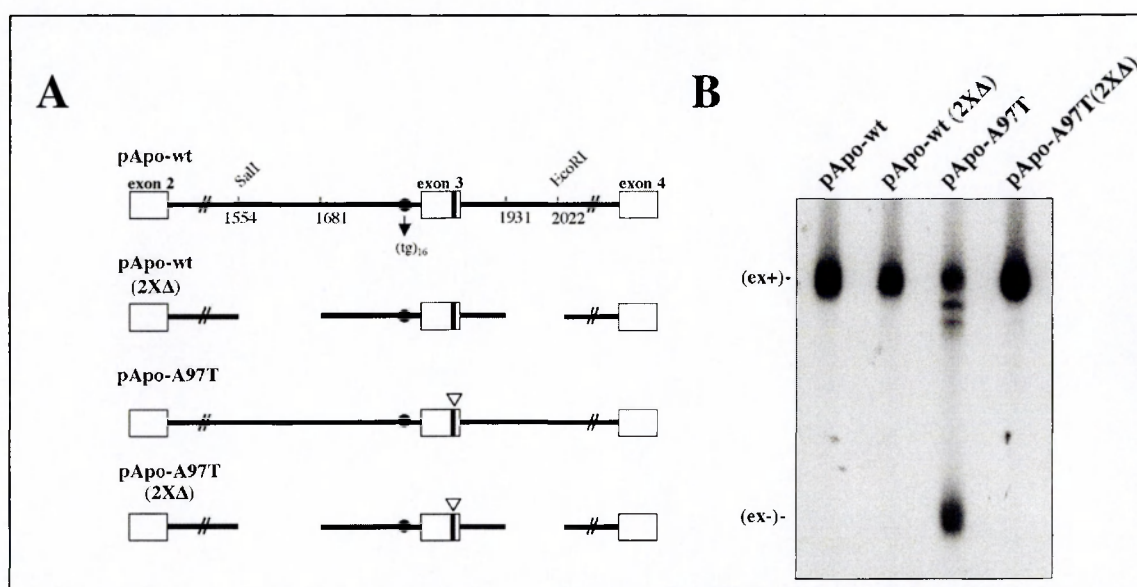


Figure 2.25. Effect of simultaneous deletions of intron 2 and 3 on Apo AII exon 3 inclusion. A) Constructs pApo-wt, pApo-wt (2X Δ), pApo-A97T and pApo-A97T (2X Δ) with the intact introns (the first two) and with deletions of 127 and 91 nt in intron 2 and intron 3, respectively. B) Splicing pattern analysis by denaturing gel electrophoresis of the transfected constructs in Hep 3B cells by RT-PCR. The effect of A97T mutation disappears when deletions are introduced in both introns.

At this point, we designed a series of pApo-wt and pApo-A97T mutants carrying either the $\Delta 1554-1681$ deletion within intron 2 or the $\Delta 1931-2022$ deletion within intron 3 to map and characterize the cis-acting elements within each intron (Fig.2.26. The single deletions affect differently the exon 3 processing. The $\Delta 1554-1681$ deletion alone neutralized the effects of the ESE mutation (Fig.2.26B, lanes 1 and 2), whereas the $\Delta 1931-2022$ deletion alone caused exon

3 skipping even in presence of the wild type ESE (Fig.2.26B, lanes 3 and 4) indicating a complex interplay between intronic and exonic sequences during the splicing process.

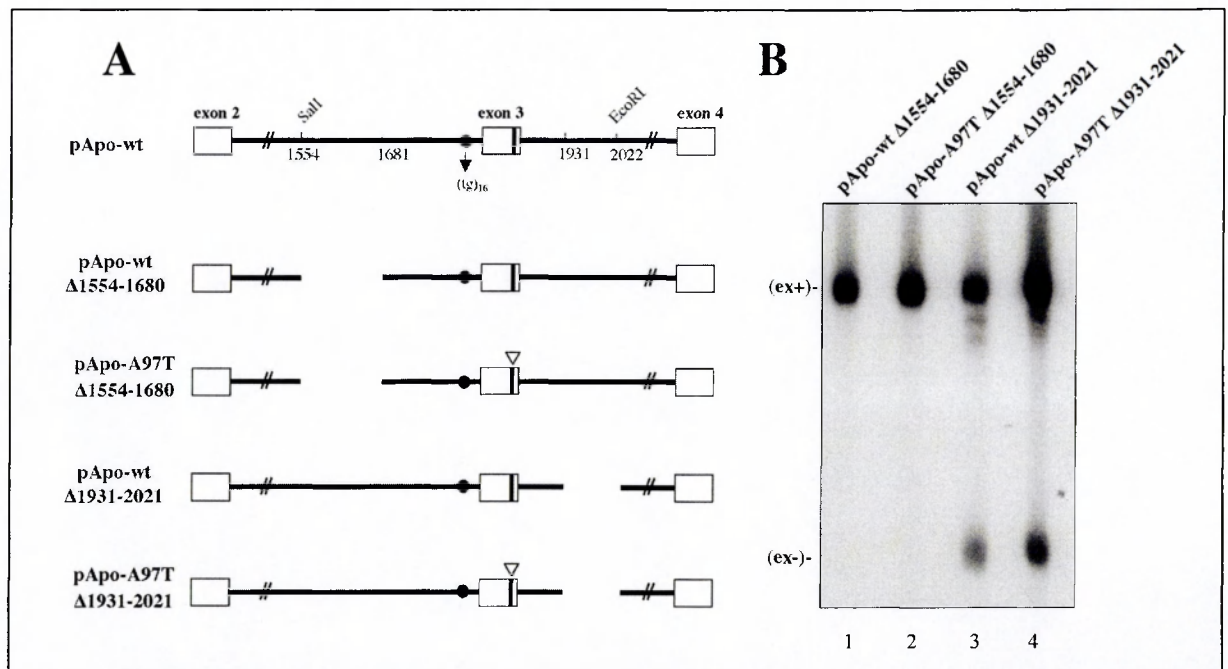


Figure 2.26. Effect of single deletion either of intron 2 or intron 3 on Apo AII exon 3 splicing. A) Constructs pApo-wt Δ1554-1681, pApo-A97T Δ1554-1681 and pApo-wt Δ1931-2022, pApo-A97T Δ1931-2022 with the deletion of intron 2 and intron 3, respectively. B) Splicing pattern analysis by denaturing gel electrophoresis of the transfected constructs in Hep 3B cells by RT-PCR. Δ1554-1681 deletion neutralizes the effects of A97T mutation whereas Δ1931-2022 deletion causes exon 3 skipping even in presence of the wild type ESE.

In order to map more finely the region responsible for these effects we first decided to add progressively the rest of the intron 2 in order to restore the whole sequence. Figure 2.27 shows that the effect of the A97T mutation within the ApoAII exon 3 ESE was recovered when the 20 bp were added (lane 6, fig 2.27). The same is observed if the next 20 bp are added (lane 4 Fig 2.27).

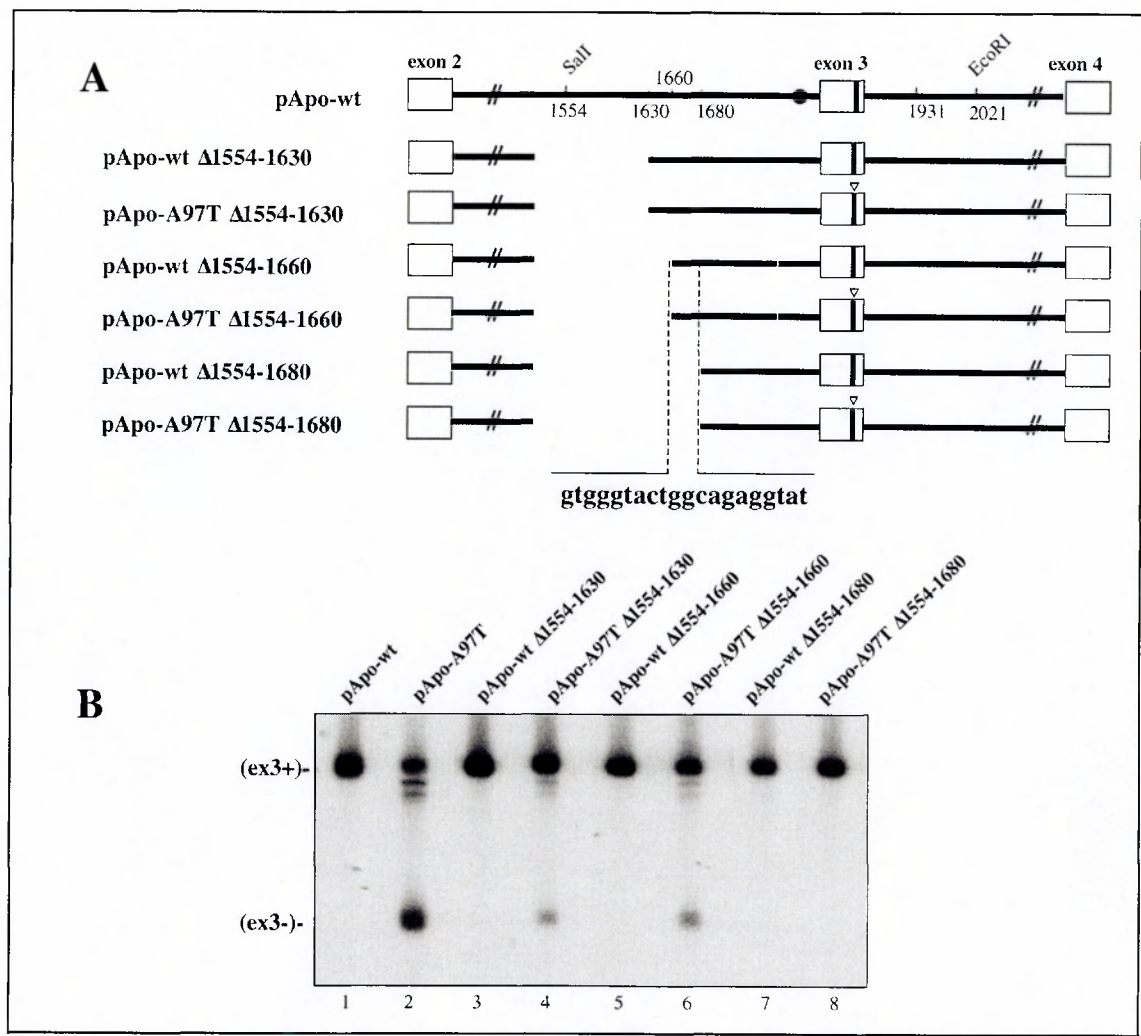


Figure 2.27. Effect of restoring intron 2 sequence on Apo AII exon 3 splicing. A) Scheme of the constructs showing the partial restoring of the intron 2 sequence. The 20 bp sequence able to restore the effect of the A97T mutation in the exon 3 ESE is indicated. B) Splicing pattern analysis by denaturing gel electrophoresis of the transfected constructs in Hep 3B cells by RT-PCR. Addition of the 20bp in construct pApoA97T Δ1554-1660 was able to restore the effect of the mutation A97T in exon 3 ESE.

The intron 3 element was also subjected to further analysis and our interest focused on a 26 bp polypurine rich region between nucleotides 1941-1967. Since many exonic and intronic splicing enhancers are purine rich, we tested the effect of a deletion covering only the polypurine tract. Figure 2.28 shows that the deletion of this tract continue to cause exon 3 skipping even when the exon 3 ESE is intact.

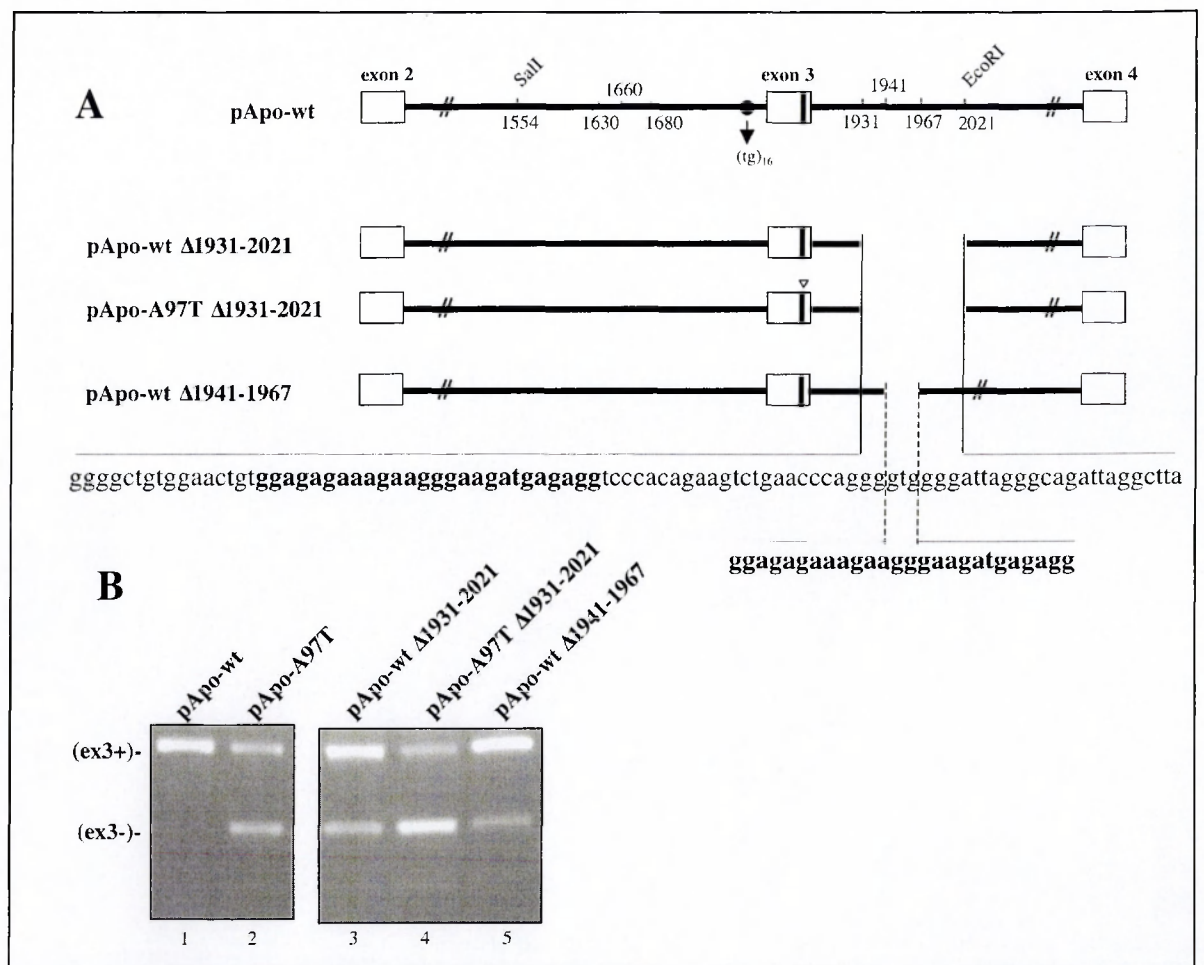


Figure 2.28. Effect of poly purine sequence deletion in intron 3 on Apo AII exon 3 splicing. A) Constructs used for transfection. The 26bp poly purine rich sequence deleted in the construct pApo-wtΔ1941-1967 in intron 3 is indicated. B) Splicing pattern analysis by denaturing gel electrophoresis of the transfected constructs in Hep 3B cells by RT-PCR. Deletion Δ1941-1967 in intron 3 causes exon 3 skipping even though the ESE in exon 3 is intact.

These results clearly indicate the presence of different accessory elements within both flanking introns. The effects of the Δ1554-1681 deletion might depend on the removal of a silencer-like sequence but, in principle, we cannot exclude that it might depend on the increase of the splicing efficiency due to the shortening of intron 2. Conversely, the effects of the Δ1941-1967 deletion suggest the presence of an intronic splicing enhancer-like sequence within that region.

3. DISCUSSION

The splicing of nuclear pre-mRNA is a highly regulated process where introns are accurately removed to produce mature mRNA. The selection of the splicing sites in the alternative and constitutive splicing is determined by complex RNA-RNA, RNA-proteins and protein-protein interactions. This thesis has been focused on the mapping of cis-acting elements and the characterization of trans-acting factors involved in the recognition of human ApoAII exon 3 a peculiar exon with an unusual 3' splice site.

An Evolutionary point of view

At the genomic level, the structure of the ApoAII gene in different species is substantially well conserved. At the mRNA level the degree of homology interspecies ranges between a minimum of 70% to a maximum of 83.2% (Fig. 3.1).

Organism-Gene	Organism-Gene	Percent ID
M.musculus -ApoAII	R.norvegicus - ApoAII	83.2
M.musculus -ApoAII	S.scrofa - Ss.3703	80.0
M.musculus -ApoAII	H.sapiens - ApoAII	76.5
M.musculus -ApoAII	B.taurus - Bt.10573	71.6
H.sapiens -ApoAII	S.scrofa - Ss.3703	83.0
R.norvegicus -ApoAII	S.scrofa - Ss.3703	78.8
R.norvegicus -ApoAII	H.sapiens - ApoAII	76.7
S.scrofa -Ss.3703	B.taurus - Bt.10573	70.0

Figure 3.1 Comparison of ApoAII genes identity among different species. The nucleotide sequence comparisons used in determining homology are listed. The pairs represent reciprocal best hits; each alignment is the best one for both organisms. The percent ID represents identity over an aligned region.

The second intron of the human ApoAII gene contains a polymorphic microsatellite sequence consisting of repetitive (GT) motifs. Recent population studies identified at least 10 different alleles of this polymorphism, ranging from 12 to 21 (GT)_n repeats (Brousseau et al., 2002).

Figure 3.2 shows a comparison of ApoAII exon 3 and the flanking IVS2 and IVS3 sequence in different species. A high level of homology of both introns and exon 3 is observed among all species except for rat, where the high homology is just limited to the exon 3.

In fact, the size of exon 3 is conserved in all species (i.e. 133bp), and the dimensions of both intron 2 and intron 3 are similar in Human, Macaca, Chimpanzee and Mouse. The dimension of rat introns is instead twice that of the other species (Fig.3.2 A). It can be seen that the intron/exon junctions share a high degree of similarity, and the splice donor-acceptor sites are identical in all species (Fig.3.2 B).

Interestingly, the (GT) tract within intron 2 is present in all Primates but not in Rodents. In Human and in Chimpanzee there are a similar number of repeats in a row (i.e. 16) whereas in *Macaca fascicularis* there are just 7 repeats. Conversely, both Mouse and Rat present 5 (GT) repeats clustered in the 3' end of intron 2 but there is no continuous (GT)_n sequence.

The high degree of ApoAII gene conservation among Primates and Rodents at the exon level and at the intron/exon junctions suggests that this gene is a result of an ancient event typical of constitutive exons. These observations are consistent with recent studies on the comparison of alternative splicing in human and mouse, which found that alternatively spliced exons are mostly not conserved and thus are the product of recent exon creation or loss events (Modrek and Lee, 2003).

Microsatellite sequences, such as (GT)_n repeats, are probably one of the best sources of multiallelic polymorphisms in the human genome (Tautz and Renz, 1984; Weber and May, 1989). Although the function of these repeats is unknown, it has been suggested that they can

act as hot spots for recombination (Hellman et al., 1988) or be involved in gene regulation (Wittig et al., 1992).

It is also possible that in the primates the GT tract is perhaps related to recent transposon activity and not to any selective improvement of the gene expression efficiency.

A

	IVS2	Exon 3	IVS3
Rat	643	133	587
Mouse	299	133	343
Macaca	273	133	394
Human	296	133	397
Chimpanzee	293	133	395

B

Rat	tgtggagtggtgaggggagatggggattcattgctggcaatccaaatacaa
Mouse	tgtggggtgtgaggggagaccgggattcattgctggcagtctacatgcag
Macaca	ctggcagaggtatggcaagaggccaagattcattgctgtggaccagctg
Homo	ggccaggattcactgctgtggaccagctgaaaagagtgtgtgtgtgtgt
Chimp	ggccaggattcattgctgtggaccagctgaaaagagtgtgtgtgtgtgt

Rat	agactgtgatattgtgtttgtgtag GAGCTTTGGTTCCGAGACAGGCAGCG
Mouse	aggctgtgatctatgtttgtgtag GAGCTTTGGTTAAGAGACAGGCGGAC
Macaca	aaaagagtgtgtgtgtgtgtgcag GAGCTTTGGTTCCGAGACAGGCAGAG
Homo	gtgtgtgtgtgtgtgtgtgtgtgtgcag GAGCTTTGGTTCCGAGACAGGCAAAG
Chimp	gtgtgtgtgtgtgtgtgtgtgtgtgcag GAGCTTTGGTTCCGAGACAGGCAAAG

Rat	GAGACGGATGTGCAGACCCTGTTTCAGCCAGTATCTTCAGAGTTTAAGTGA
Mouse	GGACAGGATATGCAGAGCCTGTTCACTCAGTACTTTCAGAGCATGACTGA
Macaca	GAGCCAAGTGTGGAGAGCCTGGTTTCTCAGTACTTCCAGACCGTGACTGA
Homo	GAGCCATGTGTGGAGAGCCTGGTTTCTCAGTACTTCCAGACCGTGACTGA
Chimp	GAGCCATGTGTGGACAACCTGGTTTCTCAGTACTTCCAGACCGTGACTGA

Rat	CTATGGCAAGGATTTGATGGAGAAGGCCAGCCCTCAGAGATTCAGAACC
Mouse	ATATGGCAAAGATTTGGTGGAGAAGGCCAAGACCTCAGAGATTCAGAGCC
Macaca	CTACGGCAAGGACCTGATGGAGAAGGTCAAGAGCCCAGAGCTTCAGGCC
Homo	CTATGGCAAGGACCTGATGGAGAAGGTCAAGAGCCCAGAGCTTCAGGCCG
Chimp	CTATGGCAAGGACCTGATGGAGAAGGTCAAGAGCCCAGAGCTTCAGGCCG

Rat	AAGCCAA gtaagtctctctgggagagagctcaaagggctacaggactgct
Mouse	AGGCCAA gtaagtctctctgggagagaactcaaagggctacaggactgcttc
Macaca	AGGCCAA gtaagtctaagggcaaggagttcaggggctgtggaactgtgga
Homo	AGGCCAA gtaagtctcagggcaaggggttcaggggctgtggaactgtgga
Chimp	AGGCCAA gtaagtctcagagcaaggagttcaggggctgtggaactgtgga

Rat	ttctttttttttttctttttctttttcttttttttttttttaagatttattt
Mouse	agggagggagggaggtctcccaggaatcttaactcaggggttggggtag
Macaca	aaggaagaaggaagatgagaggtcccacagaagactgaacccaggggtg
Homo	gagaaagaaggaagatgagaggtcccacagaagtctgaacccaggggtg
Chimp	gagaaagaaggaagatgagaggtcccacagaagtctgaacccaggggtg

Figure 3.2 Comparison of ApoAII exon 3 in different species (Rat, Human, Mouse, Macaca fascicularia and Chimpanzee). A) comparison of ApoAII exon 3 and flanking introns size in different species (Rat, Human, Mouse, Macaca fascicularia and Chimpanzee). B) alignment of ApoAII exon 3 and flanking introns size in different species, red uppercase letters indicate exon 3 sequence. Lowercase letters indicate intron 2 and intron 3 sequences.

Correlations among different sequence features of ApoAII exon 3 and CFTR exon 9

Human ApoAII exon 3 and its flanking introns share only partially the canonical features of splicing sites. In fact, the 5' splice junction of intron 3 is well defined and resembles the consensus sequences (Lopez, 1998), whereas the 3' splice junction of intron 2 contains instead of a classical polypyrimidine tract a polymorphic sequence made of alternating pyrimidines and purines.

Regarding the branch point, we focused the attention on a putative site placed at 44 bases upstream of the 3' splice site and conserved in all Primates. According to the literature data, most of the known branch points are placed in the range of 18-40 bp upstream of the 3' splice site, so we favour the hypothesis that the sequence at the position of -44 might be a functional branch point. However, a functional mapping should be carried out to confirm this hypothesis.

At least 2 other genes, human CFTR intron 8 and human cardiac alpha-actin intron 4 share with ApoAII the presence of this peculiar (GT) tract at or near the 3' splice site. The structure of human CFTR intron 8 polypyrimidine tract is composed by a combination of both (GT) and (T) polymorphic repeats. On the other hand, the architecture of the 3' splice site of the cardiac alpha-actin intron 4 is very similar to that of the ApoAII intron 2, since it also lacks a polypyrimidine tract and in its place we find a (GT) polymorphic repeat.

CFTR exon 9 undergoes alternative splicing whereas the ApoAII exon 3 and alpha-actin exon 5 are constitutively spliced. These observations, along with previous studies about the splicing of the ApoAII exon 3 (Shelley and Baralle, 1987), suggest that the alternated stretch of U and G forms a fully functional "polypyrimidine tract". Moreover, in vitro studies by Coolidge and co-workers have shown that a GU tract is able to work as a polypyrimidine tract also in other contexts (Coolidge et al., 1997).

However, this is not true for CFTR because when the (U)_n tract located immediately downstream of the stretch of GU is removed, the total exclusion of exon 9 occurred (Niksic et

al., 1999). Furthermore, our studies demonstrated that the replacement of the polypyrimidine tract of the mouse fibronectin EDA exon with the (GU)₁₆ repeats of ApoAII second intron caused skipping of this exon (Figure 2.8). Hence, the GU tract by itself does not act as a functional “polypyrimidine tract” in all the contexts.

Comparison of splice sites strength of human ApoAII exon 3 and human CFTR exon 9 has outlined an apparent contradiction between the splicing behaviour of the two exons and the strength of the splice sites. In fact, both the 3’ and the 5’ splice sites of CFTR exon 9 display a good match with the consensus but exon 9 is alternatively spliced. Conversely, the match of the ApoAII exon 3 splice site with the consensus is not good. In fact, if we applied the Splice Site Prediction by Neural Network program to the 5’ and 3’ splice site of exon 3 we obtained the authentic exon 3 donor site as the best possible, whereas the score of the authentic exon 3 acceptor site was very low and was not considered by the program as a putative 3’ splice site. Nevertheless, the ApoAII exon 3 is constitutively spliced.

It is likely that in the ApoAII exon 3 and its flanking introns there are cis-acting elements different from those present in the CFTR exon 9 context that allow the constitutive inclusion of exon 3 in the mRNA. This striking contrast in splicing pattern prompted us to investigate the mechanisms underlying the constitutive splicing of ApoAII exon 3 and, in particular, to characterize the cis-acting elements and the trans-acting factors involved in ApoAII exon 3 definition to assure its constitutive splicing.

Influence of the 3' splice site on splicing of ApoAII exon 3

Taking advantage of the compact size of the human ApoAII gene, we were able to create an “in vivo” expression system that includes the whole 1173bp-promoter, the 1226bp-coding region encompassing 4 exons and 3 introns and the 320bp-3'-UTR where the polyadenylation signal is located. In this way, it was possible to study the exon 3 splicing within its original context. This represents a clear advantage in comparison with the minigene trap system, widely used to study other splicing models, where only one exon and its flanking regions are cloned in an heterologous gene context. Thus the splicing efficiency of a specific exon might be influenced not only by the elements under investigation, but also by elements that belong to the heterologous system. The ApoAII expression system that we generated instead, contains all the elements necessary for its transcription and RNA processing that are present in the endogenous ApoAII gene.

The first approach toward the characterization of the cis-acting elements involved in ApoAII exon 3 definition, was to confirm the relevance of the (GT) tract in our controlled system. Both the deletion or the replacement of the (GT)₁₆ tract with its complementary (CA)₁₆ tract resulted in 90% exclusion of exon 3 from ApoAII mRNA (Figure 2.7).

It is well known that there is a great flexibility in the specific sequence within the polypyrimidine tract. In fact, it was shown that the introduction of purines into the polypyrimidine tract is detrimental to splicing only if the length of the tract is shortened and if there is a reduction in the number of consecutive uridine residues (Norton, 1994; Reed, 1989; Roscigno et al., 1993). Our results concerning the replacement of (GU) with (CA) repeats are consistent with previous studies by other authors indicating that Uridine and Cytidine do not appear to be functionally equivalent for the 3' splice site recognition. (Bouck et al., 1995; Fu et al., 1988; Roscigno et al., 1993).

Our *in vivo* results are coherent also with previous *in vitro* studies on tracts containing alternating pyrimidine and purine residues, which found that they are functional. Coolidge et al designed a splicing competition construct in which two polypyrimidine tracts were placed adjacent to identical branch point sequences and allowed to compete for branch point selection. The competition construct was derived from exons 2 and 3 of the rat alpha-Tropomyosin gene and their intervening sequence.

Using a “polypyrimidine tract” consisting of alternating guanosine and uridine residues (GU)₁₁, they found that it was chosen 61% of the time (Coolidge et al., 1997). The authors suggested that uridine in every other position within a polypyrimidine tract might form an entirely functional polypyrimidine tract, though the competitive efficiency of such a tract is only slightly higher than that of the competing original polypyrimidine tract and lower than seven of the other 13 tracts tested in this study. However, it should be noted that the privileged position immediately adjacent to the exon 3 of the tested (GU)₁₁ “polypyrimidine tract”, might have led to overestimate its relative strength. In fact, a similar experiment made with a construct where the original polypyrimidine tract was duplicated as a test polypyrimidine tract showed that the acceptor splice site of exon 3 just downstream of the test tract is favoured.

On the other hand, we demonstrated that the replacement of the authentic polypyrimidine tract of the EDA exon with the ApoAII (GU)₁₆ “polypyrimidine tract” is not functionally equivalent *in vivo*. Our experiments suggest that a (GU)_n “polypyrimidine tract” is weak and other *cis*-acting elements within the ApoAII gene must exist to increase the splicing efficiency of ApoAII exon 3.

As mentioned above, the peculiar (GU) repeats within the 3' splice site of ApoAII exon 3 is also present in the CFTR exon 9 3' splice site. Buratti and coworkers have demonstrated that TDP-43, a protein that originally had been shown to bind HIV-1 TAR DNA (Ou et al., 1995), binds to UG tracts (Buratti et al., 2001). We have demonstrated that TDP-43 is able to bind the

(UG)₁₆ repeats within ApoAII intron 2. Buratti et al have shown that TDP-43 is involved in the formation of the 50–52 kDa complex that assembles on the (UG)_m element at the 3' splice site of CFTR exon 9 (Buratti et al., 2001). Then, on one hand, overexpression of this protein in transfection experiments was shown to inhibit CFTR exon 9 splicing and this inhibitory role was more evident with a decrease in the number of (U)_n repeats, a result consistent with previous studies on the (UG)_m(U)_n interdependence (Niksic et al., 1999; Pagani et al., 2000). On the other hand, Buratti et al. showed that inhibition of TDP-43 expression in cultured cells by antisense PS-oligodeoxynucleotide treatment leads to an increase in exon 9 recognition. Our preliminary cotransfection experiments of the pApo-A97T construct (that presents nearly 85% of exon 3 skipping) with the same antisense phosphorotioate (PS)-oligonucleotide used by Buratti et al. strengthen the putative inhibitory effect of TDP-43 on the usage of 3' splice sites adjacent to the (UG)_n repeats. In fact, we observed a slight reduction of exon 3 skipping (60%) following the cotransfection of the TDP-43 specific antisense oligo but not with a not related antisense oligo. However, further experiments are required to elucidate the precise role that this factor is playing in the splicing of ApoAII exon 3.

5' splice site involvement in the recognition of the ApoAII exon 3

The above considerations led us to consider the relevance of the 5' splice site in the splicing of the ApoAII exon 3. In principle, the abrogation by mutagenesis of the 5' splice site is used to test the mechanism of splice site recognition. If the 5' splice site mutations results in retention of the downstream intron, this is indicative that the splicing occurs through an “intron definition” pattern. Alternatively, if the 5' splice site mutations causes skipping of the upstream exon, this indicates that the splicing occurs through an “exon definition” pattern.

In the ApoAII gene context, we found that the abrogation of the donor site of exon 3 causes the exclusion of exon 3 from mRNA. Therefore the exon 3 splicing seems to follow an “exon definition” pattern.

Other models have shown a relationship between splice site strength and splicing pattern. In particular, Talerico and Berget found that in small *Drosophila* introns with pyrimidine poor 3' splice sites, the “intron definition” pattern of splicing was usually seen, while an “exon definition” pattern was favoured when the 3' splice site was pyrimidine rich (Talerico and Berget, 1994). Thus, the ApoAII exon 3 disagrees with that model since its (GU) based “polypyrimidine tract” is weak. This suggests that the splice site strength is not sufficient to explain the “exon definition” splicing pattern, and that other parameters (such as accessory cis-acting elements) can influence the splicing model. Then, the evaluation of the splice site strength has shown that the 5' splice site of human ApoAII exon 3 is stronger than that of human CFTR exon 9 (SSPNN scores 1.0 versus 0.83). We took advantage of this difference to test the effects of lowering the strength of 5' splice site of ApoAII exon 3. Therefore, the observation that the replacement of the ApoAII exon 3 with the CFTR exon 9 donor site resulted in an increase of exon skipping suggests the possibility that the strong 5' splice site of ApoAII exon 3 might compensate for its weak 3' splice site. The possible explanation of the mechanism of compensation is given by an indirect communication between U1 snRNP and U2AF. In fact, Hoffman and Grabowski have indicated that a strong 5' splice site can compensate for a weak 3' splice site. Indeed, they found that in the case of the rat preprotachykinin pre-mRNA, the essential splicing factor U2AF65 might be targeted to a weak 3' splice site in a manner that is facilitated by U1 snRNP particles and the downstream 5' splice site. It was shown the importance of the presence of an intact 3' splice site, since U1 snRNP binding to the 5' splice site was not sufficient to target U2AF65 binding when the upstream tract was destroyed by mutation (Hoffman and Grabowski, 1992). Thus, the ApoAII

exon 3 splicing mechanism does fit with this model based on the indirect communication between splice sites across the exon.

These observations point to the existence of accessory cis-acting elements placed within the exon that might work as a bridge between the exon borders.

ApoAII exon 3 sequences and splicing regulation

Studies on regulated alternative splicing have identified exonic cis-acting elements, referred to as exonic splicing enhancers (ESEs) that facilitate the process of exon definition. These elements are capable of activating weak splice sites in adjacent introns. However, it is now clear that ESEs are not only components of regulated exons but also of constitutively spliced exons (Mayeda et al., 1999; Schaal and Maniatis, 1999a).

Two classes of exonic splicing enhancers have been reported. The AC-rich elements (ACE) are a class of ESE identified by in vivo selection experiments and were found to stimulate splicing both in vivo and in vitro (Coulter et al., 1997). AC-rich ESEs have been shown to be involved in the regulated splicing of both viral and cellular genes (Coulter et al., 1997; Gersappe and Pintel, 1999). The mechanism of action of AC-rich ESEs remains largely unknown, although these elements have been proposed to function in a manner similar to that of the purine-rich ESEs.

The purine-rich ESEs are the most common and are usually located downstream of a suboptimal 3' splice site. Through interactions with serine-arginine-rich (SR) proteins, purine-rich ESEs recruit U2AF to suboptimal 3' splice sites and stimulate spliceosome assembly (Graveley et al., 1998; Lavigne et al., 1993; Nagel et al., 1998; Ramchatesingh et al., 1995; Sun et al., 1993; Tanaka et al., 1994; Zheng et al., 1997; Zuo and Maniatis, 1996). Purine-rich ESEs can also suppress splicing of a pre-mRNA when they are located in a regulated intron

(Gallego et al., 1997; Kanopka et al., 1996; Pagani et al., 2000). Thus, a purine-rich ESE can function as a splicing enhancer or a splicing suppressor depending on its location.

The accessory cis acting elements within the exon 3 were initially investigated by the progressive replacement of ApoAII exon 3 with CFTR exon 9 sequences. This approach was favoured to a deletion-based analysis, because our good knowledge of CFTR exon elements (Pagani et al., 2003) and to preserve the spatial arrangements of the splice sites. Indeed previous mapping analysis of human CFTR exon 9 for intraexonic regulatory elements carried out in our laboratory were based on the creation of human /mouse exon 9 hybrids because of the different exon 9 splicing outcome in human and mouse. The use of murine ApoAII gene was not possible in this case because murine and human ApoAII exon 3 have an identical splicing pattern.

The progressive replacement of ApoAII by CFTR sequences lead to the identification of a region between nucleotides 87 to 113 of human ApoAII exon 3 as essential for its inclusion in mRNA. Further mapping studies by overlapping deletions in that region permitted to identify a 9nt-core region with enhancer-like activity, located in the region spanning from nucleotide 91 to 99 of exon 3. Considering the base composition of this 9nt-core region (TGATGGAGA) this sequence would belong to a purine-rich ESE category.

In order to map more finely the 9nt-core region, we then evaluated the effect of point mutations within the core region on the exon 3 definition. Our results showed that single nucleotide substitutions may have a profound effect on the splicing efficiency inducing skipping. This peculiar sensitivity to splicing variations may be related to the weak definition of the ApoAII exon 3. This observation strengthen the hypothesis of the possible relationships between the definition of the exon 3 mediated by the strength of the 3' splice site and the exonic regulatory structures affected by the site-directed mutants. Another possibility that we cannot rule out is that the mutations A97T, A99T and G92C (fig.2.15) create a silencer like

sequence, rather than an enhancer. However the 9nt putative enhancer sequence (TGATGGAGA), contains the motif GGAGA, which has been found in many other purine-rich enhancer elements and has been recovered from a random screen (Schaal and Maniatis, 1999b). Two other studies based on *in vitro* selection had previously identified consensus sequences for splicing enhancers that depend on specific SR proteins (Liu et al., 2000; Liu et al., 1998). Following these considerations, we decided to investigate the possible trans-acting factors able to bind specifically the ApoAII exon 3 ESE.

Trans acting factors and ApoAII exon 3 splicing

It has been suggested that exonic splicing controlling elements affected by point mutations contain consensus sequences derived from SR-specific score matrices that mediate exon splicing efficiency (Cartegni L, 2002; Liu et al., 2001; Liu et al., 2000; Liu et al., 1998). In order to identify the trans acting factors able to bind the putative enhancer sequence, we tested the presence of SR protein binding sites within both wild type and A97T ApoAII exon 3 using the four available SF2/ASF, SC35, SRp40, and SRp55 motif-scoring matrices.

We found that the wild type ApoAII exon 3 contains 8 high score motifs for ASF/SF2, six for SC35, seven for SRp40, and four for SRp55. Interestingly, we found one ASF/SF2 potential site starting at nucleotide 90 (CTGATGG), that overlaps with the putative enhancer mapped functionally between nucleotides 91-97. However, no differences in the calculated matrix scores for the four SR proteins were apparent for the substitutions (A97T, G92C and A99T) that were found to affect “in vivo” the ApoAII exon 3 splicing pattern.

This is also in contrast with the SR protein binding data. Indeed whilst the wild type (wt) and A97T ESE sequences have both a high score for ASF/SF2, our binding experiments show an extremely clear difference with good binding to the wt ESE that is abolished by the A97T mutation.

The SR score matrices have been previously found to show a relationship with the splicing phenotype in other gene systems (Cartegni and Krainer, 2002; Liu et al., 2001). However, Pagani and co-workers when tried to correlate the SR protein matrices scores calculated for human CFTR exon 9 with the results of transfections of mutants generated by site-directed mutagenesis of exon 9, found a lack of general predictive value of the SR matrices for the identification of *cis*-acting elements in this context (Pagani et al., 2003). Another study of ESE analysis scored binding potentials for the different SR proteins in the dystrophin gene and could not find a relationship for the observed splicing patterns. In fact, in many cases exons

that were found to be skipped in at least one alternative transcript displayed high SR protein binding potentials (Sironi et al., 2002).

Therefore, the “in silico” cannot be taken as a general prediction of SR binding sites and in particular does not seem to apply to ApoAII exon 3 *cis*-acting elements as observed also by Pagani et. al. and Sironi et. al. for other contexts. Hence, their weight in splicing regulation cannot be taken into consideration without a proper evaluation that should necessarily include binding and functional splicing assays in each specific system.

Considering that SR proteins others than those tested by matrix scores might be involved in the splicing of ApoAII exon 3, we tested the effects of overexpression of different factors on the splicing efficiency of this exon. It was necessary to test positive effects on a construct different from the wild type that already displays almost 100% inclusion. Therefore the 1866-86 deletion that produces 55% of exon exclusion was used.

Among the tested factors (SRp20, SRp55, SRp75, ASF/SF2, SC35 and hnRNP A1), overexpression of SRp55, SRp75, ASF/SF2 and SC35 resulted in a slight increase of exon 3 inclusion, whereas SRp20 and hnRNPA1 had a negative effect causing decreasing of exon 3 inclusion. Our results with SRp55 and SRp75 are consistent with literature data since SRp55 (Tran et al., 2003; Tran and Roesser, 2003) and SRp75 (Ramchatesingh et al., 1995; Zheng et al., 1997) are known as SR proteins binding to active enhancers.

The smaller than expected effect of ASF/SF2 and SC35 overexpression might be due to the fact that the 1866-86 deletion, occurring just 5nt downstream of the ApoAII exon 3 ESE, partially alters the RNA secondary structure of the region so interfering with the correct display of the ESE.

Concerning the negative effect on exon 3 splicing caused by SRp20 overexpression, we can speculate that might be the effect of a putative interaction with a decoy target site. This putative site could be an intron silencer that would cause the displacement of the spliceosomal

machinery from the “true” exon 3 to a “decoy exon”, so preventing the constitutive inclusion of exon 3 at a certain extent. This explanation has already been proposed by Pagani et al, in the study on the modulation of CFTR exon 9 skipping (Gallego et al., 1997; Kanopka et al., 1996; Pagani et al., 2000).

The negative effect of the overexpression of SRp20 is interesting as this factor is generally associated with a positive effect on exon inclusion (Galiana-Arnoux et al., 2003; Lou et al., 1998; Schaal and Maniatis, 1999b).

On the other hand the effect of hnRNPA1 on ApoAII exon 3 splicing is consistent with the general inhibitory effect described for this factor (Caceres et al., 1994; Mayeda and Krainer, 1992).

Binding studies by UV-crosslinking alone did not evidence clear interactions of proteins in the molecular weight range of the SR proteins of interest (SRp 20, 55 and 75). Therefore, future experiments combining UV-crosslinking with immunoprecipitation using other monoclonal antibodies (such as Mab104 or 1H4) that recognize different SR proteins might highlight putative interactions of SR proteins others than SC35 and ASF/SF2 with ApoAII exon 3 and its flanking introns.

Secondary structure

Since RNA molecules are generally sequence-dependent folded structures rather than simple strings of nucleotides, we included a secondary structure analysis of ApoAII exon 3 plus parts of its flanking introns in order to have a spatial prediction of the cis acting elements studied so far.

Current evidence concerning protein-RNA interaction suggests that proteins often make sequence-specific contacts with single-stranded residues in loops, while base-paired regions are often necessary for presentation of single-stranded nucleotides in the correct configuration. Several studies have suggested that some splicing events might be explained by structural alterations in possible higher free-energy structures of the pre-mRNA at specific enhancer or silencer elements (Balvay et al., 1993). Here in, it was hypothesized that when an alternative structure is too different from the “ground state” (lowest free-energy structure) this sequence can be differently recognized by the splicing machinery.

In order to highlight the possible relevance of the secondary structure for the ESE within the ApoAII exon 3, we elaborated this hypothesis analyzing the lowest free-energy structures of the ApoAII exon 3 wild-type (WT) using the mfold Server (<http://BiBiServ.TechFak.Uni-Bielefeld.DE/mfold/>) (Fig.3.3).

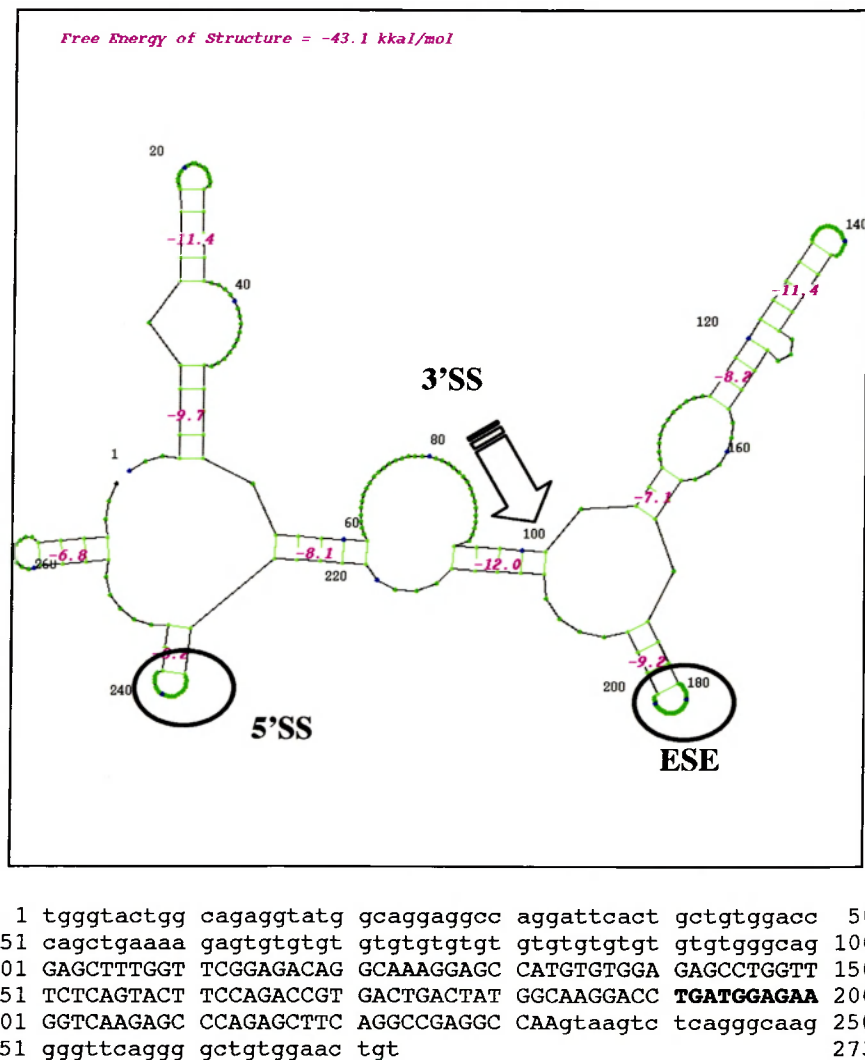


Figure 3.3. Lowest free-energy structure of the ApoAII exon 3 wild-type (WT). The structure was obtained using the Mfold Server. Black circles highlight the 5'splice site and the ESE within exon 3. The 3'splice site is indicated by the white arrow.

The computer analysis shows that both the acceptor and the donor splice site of exon 3 are included in a loop and therefore are accessible to splicing factors. Moreover, the region of the putative enhancer is also included in a stem-loop structure (nucleotides 190-200 in figure 3.3). This is consistent with a role of the 91-99 nucleotide range as an enhancer like sequence folded to form a loop structure able to interact with possible trans-acting factors (Muro et al., 1999).

The fact that both the acceptor and the donor sites are placed within a loop-structure that

should facilitate their utilization is consistent with previous studies regarding the splicing patterns of the human growth hormone transcript, in which two bases in the vicinity of the exon 3 major splice-acceptor site (B) facilitate the utilization of a competing downstream acceptor (B'). In this case it was demonstrated that these two bases function by stabilizing a specific stem-loop structure in the native transcript indicating that a specific secondary structure within the native human growth hormone transcript controls the relative utilization of two competing splice-acceptor sites with the consequent generation of two functionally distinct hormone isoforms (Estes et al., 1992).

Another example regards the alternative splicing of the calcitonin/CGRP pre-mRNA. Here, both computer analysis and structure specific RNase digestion analysis of the RNA containing the 3' splice acceptor of the calcitonin-specific exon 4, predicted that this region has the potential to form a thermodynamically stable stem-loop, acting as a crucial cis-element involved in proper splice site selection (Coleman and Roesser, 1998).

Then, our attention was focused on the ApoAII exon 3 region including the ESE. In our model, to maximize the structural information provided by the RNA molecule in an RNA-protein interface for the prediction of the ApoAII exon 3 secondary structure, the region used for the computer analysis included all the exon. The ESE is included within a hairpin loop structure available to interact with trans-acting factors as shown for other models.

In fact, different examples of accessory splicing factor binding dependent on the secondary structure of target sequence, have already been reported. It was shown that binding affinities of both human and drosophila SRp55 are strongly influenced by the structural features of the target RNA (Shi et al., 1997). Then, the binding of hnRNP A1 (its well known antagonistic factor) has been recently observed to be affected by RNA secondary structure in the removal of the second intron in the HIV-1 rev/tat pre-mRNAs (Damgaard et al., 2002).

The position of the ESE within ApoAII exon 3 is crucial

The “in vitro” splicing experiments carried out in order to investigate the possible relationships of the ApoAII exon 3 ESE with the flanking 3’ splice site did not show any difference in the splicing pattern either with the ESE wild type or with the A97T point mutation (Fig. 2.21). Although a putative influence of the PY7 minigene’s context cannot be ruled out, the fact that the ESE is located within the “last” exon in this minigene, may affect the function of the ESE, normally placed within an internal exon. In addition, our “in vivo” experiments have shown that the ApoAII exon 3 ESE works properly only when it is located within an internal exon (Fig. 2.23).

It is important to note that the splicing mechanisms of the first and the last exons are peculiar since first exons contain an m7GpppG 5’ cap structure instead of a 3’ splice site and last exons contain a poly A signal instead of a 5’ splice site.

The 5’ cap structure of a mammalian pre-mRNA is necessary for the efficient utilization of the adjacent 5’ splice site for definition of the first exon and removal of the first intron (Izaurrealde et al., 1994). Efficient recognition of the cap-proximal 5’ splice site by the U1 snRNP is facilitated by the nuclear cap binding complex (CBC) (Lewis et al., 1996). This CBC-directed definition of the first exon is believed to be one of the earliest steps in pre-mRNA recognition and appears to involve an interaction between the CBC (bound to the cap) and the U1 snRNP (which binds at the 5’ splice site). On the other hand, processing of the last exon and removal of the last intron involve interactions between splicing components at the 3’ splice site of the last exon and components of the Polyadenylation (PA) complex at the poly A signal (Lutz and Alwine, 1994; Nesic et al., 1993; Niwa and Berget, 1991; Niwa et al., 1990).

In principle, the repositioning of ApoAII exon 3 ESE within the last exon does not imply necessarily an alteration of its function. In fact, there are examples of last-exon splicing that depends on the presence of an enhancer within the last exon. Splicing of the last intron (intron

D) of the bovine growth hormone (bGH) pre-mRNA requires the presence of a downstream splicing enhancer in the last exon (Dirksen et al., 1994; Hampson et al., 1989).

Therefore, our observation that the ApoAII exon 3 ESE works properly only when it is located within an internal exon (fig. 2.23), implies the existence of other regulatory elements placed within the flanking introns.

Mapping of cis-acting elements within intronic regions flanking ApoAII exon 3

There are several splicing control elements located within introns (Smith and Valcarcel, 2000). They can be involved in recognition of the appropriate splice site or in the suppression or enhancement of certain splice site usage, particularly in alternative splicing cases. It is difficult to make a clear distinction between splicing recognition and regulation, since an element may serve as a constitutive splicing enhancer/suppressor or promote alternative splicing, depending on the expression pattern of the gene and the availability of trans-acting associated factors.

Relatively little is known about the nature of the intronic splicing regulatory elements (ISRE). Generally, they do not obey defined consensus rules and are known to consist of diverse sequences (Fairbrother and Chasin, 2000). Although some common ISRE have been observed, such as the GGG triplet (Carlo et al., 1996; McCullough and Berget, 1997; Romano et al., 2002), purine-rich elements (Hastings et al., 2001; McCullough and Schuler, 1997), or polypyrimidine tracts commonly present in the 3' intronic regions (Hastings and Krainer, 2001), most newly discovered elements tend to consist of entirely new motifs, rather than belong to known groups (Blencowe, 2000).

When we analyzed the sequence of the ApoAII intron 3 scanning for possible ISREs, we noticed the presence of three G runs between nucleotide 10 and 28 downstream of the 5' splice site of intron 3. However, our results of Gs runs mutagenesis excluded their involvement in the splicing control of ApoAII exon 3 (Fig. 2.25).

Then, to evaluate the functional significance of the human ApoAII intron 2 and 3 in modulating splicing efficiency of exon 3, intronic sequences were progressively deleted and the corresponding mutant minigenes were transiently transfected in Hep3B cells and the resulting mRNA analyzed by RTPCR.

It is known that the potential localization of ISREs is extremely variable. In fact, they are most likely present in the vicinity of a splice site but have also been found as far away as 150 bp (McCullough and Schuler, 1997; Rowen et al., 2002; Zheng et al., 2000).

We therefore considered the possible presence of regulatory elements in a range of 210nt upstream of the exon 3 acceptor site and 116nt downstream of the exon 3 donor site, as this was the cassette originally designed in order to facilitate sequence replacements within exon 3 and its flanking intronic sequences.

The deletions within introns had two opposite effects on exon 3 splicing. On one hand, the 127nt-deletion within intron 2 (Δ 1554-1681) caused full inclusion of the exon 3, regardless the point mutation A97T within the exonic splicing enhancer (fig.2.26). On the other hand, both the 90 nt-deletion in intron 3 (Δ 1931-2022) and a narrower one (Δ 1941-1967) caused exon 3 skipping even when the wild type ApoAII exon 3 ESE was present (fig.2.27 and 2.29).

Concerning the effects of the 127 nt-deletion within intron 2, it may indicate the presence of a silencer-like intronic regulatory element placed within the deleted region. However this possibility should be taken with caution because variations in the length of an intron are known to influence the efficiency of the splicing. In fact a compensatory relationship between exon and intron sizes has been previously described (Chen and Chasin, 1994; Sterner et al., 1996). These authors found that expanding small introns flanking a large exon causes exon skipping. They concluded that large exons bordered by large introns create an exon/intron architecture that is problematic for splice site recognition. Moreover, Bell and coworkers proposed a kinetic probability model in which short introns may overcome the poor

recognition of alternatively spliced exons (Bell et al., 1998). Using the murine CD44 V3 exon as a splicing model, the authors demonstrated a relationship between intron length and V3 exon exclusion. In fact, it was showed that the splicing of V3 is reduced as intron length is increased (Bell et al., 1998)

Conversely, Bai and coworkers showed that the expansion of introns can be associated with an increase of the splicing efficiency. In fact, using the adenine phosphoribosyltransferase gene model, they have shown that the expansion of intron 3 from 154 to 424nt greatly reduced exon 4 skipping when transfected into CHO cells (Bai et al., 1999). In consequence the increase of ApoAII exon 3 splicing efficiency might be due simply to shortening of the intron 2 by the 127nt-deletion or to an intronic splicing silencer present in that sequence. Future studies based on a point mutation strategy should be helpful to elucidate the possible presence of a regulatory element within ApoAII intron 2.

On the other hand, the observation that the 26nt-deletion within intron 3 cause exon 3 skipping even in presence of the wild type ESE, strongly indicates that a cis acting element that positively regulate the inclusion of exon 3 is localized in that sequence. There are several examples of intronic splicing enhancers (ISEs) that affect splicing of exons. The high contents of purines also might be indicative of a regulatory sequence since it seems to be the hallmark of these elements (Blencowe, 2000). The fact that this region might be regulating positively exon 3 inclusion is also reinforced by preliminary UV crosslinking analysis (data not shown) indicating four proteins specifically bound to the transcript with the intact ISE and absent when that sequence was mutagenized. This suggests that some trans acting factors might be acting through the binding to the ISE in intron 3 in order to positively regulate the inclusion of the exon 3. Nevertheless, further experiments also in this case, are necessary in order to map finely the 26nt-region by point mutation and to identify the trans-acting factors able to bind the putative ISE.

Models of human ApoAII exon 3 constitutive splicing

On the basis of our data, we can try to depict a model to explain the constitutive splicing of the human ApoAII exon 3. Different lines of evidence indicate the peculiarity of the exon 3 splicing mechanism. Firstly, a series of studies suggest that a (GU) repeats is unable to directly bind the constitutive splicing factor U2AF65 (Singh et al., 1995; Zamore and Green, 1991; Zamore et al., 1992). Moreover, our results demonstrated that the “polypyrimidine tract” of ApoAII intron 2 based on (GU) repeats are able to interact with TDP-43 and that this protein affects negatively ApoAII exon 3 inclusion, as also shown in the CFTR context. Secondly, we have identified one enhancer-like sequence localized within exon 3 and at least one intronic enhancer-like sequence localized within intron 3. Thirdly, the analysis of the introns removal order suggests that intron 2 is the last one to be processed. Figure 3.4 summarizes the cis-acting elements and trans acting factors identified during this study along with their possible effects on the ApoAII exon 3 splicing.

On the basis of our data we propose that the weak polypyrimidine tract of intron 2 might be unable to recruit directly the constitutive splicing factor U2AF65. In fact, the interaction of TDP-43 with the (GU) repeats might hamper the binding of U2AF65 with the 3' splice site of intron 2. Moreover, the presence of a strong 5' splice site is not sufficient to compensate the weak 3' splice site. Therefore, both the exonic and the intronic enhancers might have a pivotal role in definition of exon 3. The binding of ASF/SF2 (and/or SC35) to the enhancer within exon 3 and the interaction of yet unknown splicing factors to the regulatory element within intron 3 might be crucial to recruit the constitutive splicing factors U2AF35 and U2AF65 essential to define the 3' splice site of intron 2. In this way, a bridge-like structure among components bound to the 5' and 3' splice sites would be formed to define exon 3. A large body of evidences indicates that U2AF35 is required for constitutive splicing and also works as a mediator of enhancer-dependent splicing. In vitro protein-RNA interaction studies with pre-

mRNAs containing either a constitutive or regulated splicing enhancer have shown that U2AF35 directly mediates interactions between U2AF65 and proteins bound to the enhancers (Zuo and Maniatis, 1996). Thus, U2AF35 recruits U2AF65 acting as a bridge between this molecule and the enhancer complex. Without the enhancer, U2AF65 should not bind, due to the weakness of the polypyrimidine tract.

Interestingly, the observation that intron 2 seems to be the last one to be processed led us to hypothesize that this delay might be necessary to permit, through an unknown mechanism that might involve the factors interacting with the regulatory elements placed within exon 3 and its flanking introns, the displacement of TDP-43 from the (UG) repeats within the 3' splice site of intron 2, so permitting the processing of this intron.

In conclusion, in the mammalian genomes there are abundant traces of recombination events via transposons or retroviruses that resulted in substantial changes in specific regions of the genomes (Kazazian, 1998; Kazazian and Moran, 1998). As mentioned above, the polymorphic (GT) tract within ApoAII intron 2 is present in all Primates but not in Rodents suggesting that exon 3 and its flanking introns are the result of ancient insertion- and rearrangement- events, similarly to what happened for CFTR intron 8. These events would have placed the new polymorphic cis-acting element in proximity of the 3' splice junction of ApoAII intron 2 which interferes with efficient splicing of exon 3 and may have activated an evolutionary selection of different mechanism of splicing in order to preserve the constitutive inclusion of ApoAII exon 3.

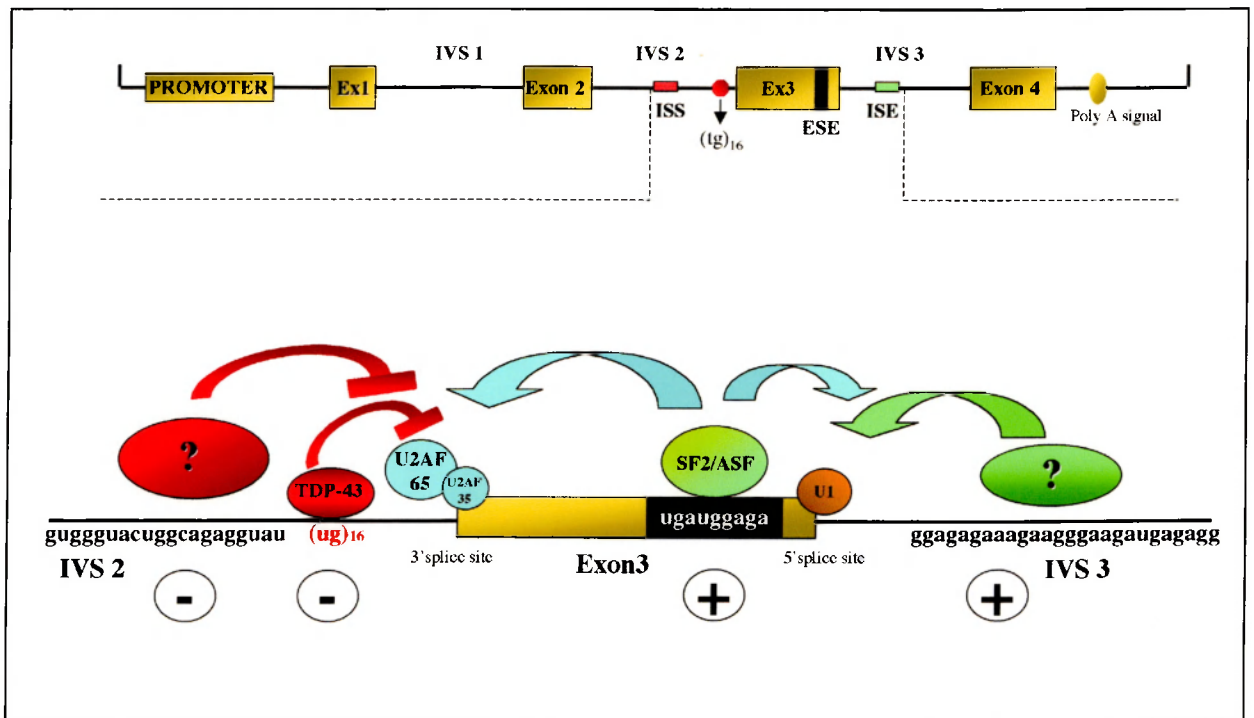


Figure 3.4 Scheme of the cis-acting elements and trans acting factors involved in the human ApoAII exon 3 definition. TDP-43 was found to interact with the (UG)₁₆ in the 3'splice site of ApoAII intron 2. ASF/SF2 and SC35 were found to bind the ESE within the exon 3. The sequences of the putative cis-acting elements localized within the intron 2 and 3 are indicated. The positive or negative effects of each cis-acting element on the exon 3 splicing is shown

Future directions

The effect observed upon deletions of the 1554-1681 region within intron 2 might be indicative of the presence of an intronic splicing silencer. However an intron-size effect due to the shortening of the introns cannot be excluded. Therefore, a strategy involving point mutations in that region without altering the size of the introns might be used in order to confirm the presence of the putative cis-acting elements within the intron 2.

Then, the preliminary results of the UV crosslinking experiments with the wild type or the mutated 26nt polypurine rich region within intron 3 could be confirmed through UV crosslinking competition assays. If is possible to highlight a specific protein able to bind the 1941-1967 region within intron 3, RNA-affinity purification and microsequencing by mass spectrometry might be carried out for the identification of such a protein.

In order to investigate more in detail the role of TDP-43 in the ApoAII exon 3 splicing, an *in vitro* splicing assay where ApoAII Δ 1866-86 transcripts should be incubated with TDP-43 immuno-depleted HeLa nuclear extracts might be set up. In such a way, the effects of TDP-43 depletion and complementation on the splicing efficiency of ApoAII exon 3 might be tested.

It was speculated that the interaction of TDP-43 to the UG repeats close to the 3' splice site might interfere with the recruitment of U2AF65 that is essential for the recognition of the acceptor splice site. Thus, an UV-crosslinking assay followed by immunoprecipitation experiments using antibodies anti-U2AF65 might clarify whether this factor is directly or indirectly recruited to the ApoAII intron 2 polypyrimidine tract.

A new tool called CLIP (ultraviolet cross-linking and immunoprecipitation) to investigate the *in vivo* RNA-protein interaction has been recently described. This technique is based on *in vivo* UV-irradiation to form covalent bonds between proteins and RNAs that are in direct contact (within angstroms); the relatively low efficiency of this reaction is compensated by

PCR amplification. The covalent binding allows not only to obtain highly purified protein-RNA complexes, but also to partially digest the RNA while retaining the core element involved in protein binding (generally 60- to 100-nucleotide). The protein component is removed from the complex with proteinase K, and the RNA cloned with the use of linker ligation and reverse transcription RTPCR, in that way, both the identification of bound RNA species and location of the binding site may be carried out [Ule, 2003 #1656]. In order to gain further insight into the role this factor is playing in the ApoAII exon 3 splicing mechanism, a similar procedure might be used to in vivo confirm the binding of TDP-43 with the (GU)_n repeats in the ApoAII context.

Finally, both the positive effects shown by SRp55 and SRp75 on exon 3 inclusion and the negative effect of SRp20 suggested the possibility that also these splicing factors might be playing a role in the splicing mechanism of exon 3. Also in this case, UV-crosslinking assays combined with immunoprecipitation using other monoclonal antibodies that recognize different SR proteins (such as Mab104 or 1H4) might be useful to map the sequences of ApoAII exon 3 and its flanking introns interacting with these SR proteins.

4. MATERIALS AND METHODS

Chemical reagents

General chemicals were purchased from Sigma Chemical Co., Merck, Gibco BRL, Boehringer Mannheim, Carlo Erba and Serva.

Standard solutions

All solutions are identified in the text except for the following:

- a) TE: 10 mM Tris-HCl (pH 7.4), 1 mM EDTA (pH 7.4)
- b) PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4
- c) 10X TBE: 108 g/l Tris, 55 g/l Boric acid, 9.5 g/l EDTA
- d) 6X DNA sample buffer: 0.25 % w/v bromophenol blue, 0.25 % w/v xylene cyanol FF, 30 % v/v glycerol in H₂O.
- e) 10X protein sample buffer: 20 % w/v SDS, 1 M DTT, 0.63 M Tris-HCl (pH 7), 0.2 % w/v bromophenol blue, 20 % v/v glycerol, 10 mM EDTA (pH 7).

Enzymes

Restriction enzymes were from New England Biolabs, Inc. DNA modifying enzymes such as Taq Polymerase, DNase I RNase free, and T4 DNA ligase were obtained from Roche Diagnostic. Klenow fragment of *E. coli* DNA polymerase I and T4 polynucleotide Kinase were from New England Biolabs, Inc. RNase A was purchased from Sigma Chemicals Ltd. A 10 mg/ml solution of RNase A was prepared in sterile water and boiled for 10 minutes to destroy trace amounts of DNase activity. All enzymes were used following manufacturer's instructions.

Synthetic oligonucleotides

Synthetic DNA oligonucleotides were purchased from Sigma-Genosys Ltd.

Radioactive isotopes

Radioactive α -³²P-UTP were supplied by Amersham U.K. Ltd.

Bacterial culture

The *E. Coli* K12 strain DH5 α was transformed with the plasmids described in this study and used for their amplification. Plasmids were maintained in the short term as single colonies on agar plates at 4 °C but for long term storage they were kept on glycerol stocks made by adding sterile glycerol to a final 30% v/v concentration to liquid bacterial cultures. Glycerol stocks were stored at –80°C. When necessary, from the glycerol stocks an overnight culture of bacteria was grown in Luria-Bertani medium [LB medium: per litre: 10 g Difco Bactotryptone, 5 g Oxoid yeast extract, 10 g NaCl, (pH 7.5)]. Bacterial growth media were sterilized before use by autoclaving. When appropriate, ampicillin was added to the media at a final concentration of 200 μ g /ml.

Cell culture

The cell line used for transfection and cotransfection experiments were Hep3B: human, hepatocellular carcinoma fibroblast-like, HeLa and the mouse fibroblast NIH 3T3 cell line.

DNA preparation

Small scale preparation of plasmid DNA from bacterial cultures

Rapid purification of small amounts of recombinant plasmid DNA was performed with the method previously described by Sambrook (Sambrook et al., 1989). Briefly, alkaline lysis of

recombinant bacteria was performed by resuspending the bacterial pellet in 200 µl of ddH₂O; 150 µl of solution II (0.2 M NaOH, 1 % w/v SDS) were then added and the contents mixed by inversion. 250µl of solution III (3 M potassium acetate pH 5.2) were then added and the contents mixed by inversion. The bacterial lysate was then centrifuged in an Eppendorf microcentrifuge at maximum speed and the supernatant transferred to a new tube. An equal volume of 1:1 v/v phenol:chloroform solution was added to the supernatant. The tube was then vortexed and centrifuged as above. The aqueous phase containing the DNA was transferred to a new tube. An equal volume of chloroform was added to the supernatant. The tube was then vortexed and centrifuged as above. The aqueous phase containing the DNA was then recovered and the DNA pelleted by ethanol precipitation. The final pellet was resuspended in 50 µl of ddH₂O and 5 µl of such preparation were routinely taken for analysis by restriction enzyme digests.

Large scale preparations of plasmid DNA from bacterial cultures

For large-scale preparations of plasmid DNA that was necessary for the transfection experiments, JetStar purification kit (Genomed) was used according to the manufacturer's instructions. In order to get a good amount of plasmid, we left a 30 ml of overnight bacterial culture using TB medium.

Preparation of genomic DNA

Peripheral blood (5 ml) was treated with 20 ml of Lysis Buffer (Sucrose 0.32 M, Tris-HCl 10 mM pH 7.5, MgCl₂ 5 mM, Triton X-100 1%) and incubated on ice for 15 min. The mix was centrifuged 10 min at 1000xg at 4 °C. The pellet was washed three times with Fisis Buffer (NaCl 0.075 M, EDTA 0.025 M pH 8) and resuspended in 1 ml of Resuspension Buffer (Tris-HCl 10 mM pH 7.5-8, NaCl 0.4 M, EDTA 2 mM). 200 µl of SDS 10% were added and the

mixture was incubated at 37 °C overnight to facilitate the extraction. The following day 600 µl of NaCl saturated water solution was added and the samples were mixed for 15 seconds to allow the proteins precipitation. The mix was centrifuged 15 min at 1500g to pellet the proteins. The supernatant was recovered and the centrifugation repeated a second time. Finally, 1 volume of isopropanol was added to the supernatant to precipitate the DNA. Precipitated DNA was removed with a Pasteur, washed in 1ml of 70% ethanol and resuspended in 500 µl of water. The DNA was then checked by electrophoresis on a 0.8% w/v agarose gel, quantified by spectrophotometer and stored at 4 °C until use for PCRs.

RNA preparation from cultured cells

Cultured cells were washed with PBS and then RNAwiz provided from Ambion inc. was added. Then, chloroform extraction was performed. Supernatant was precipitated with isopropanol. The pellet was resuspended in 100 µl of ddH₂O and digested with 1U of DNase RNase free. The mix was incubated at 37 °C for 30 minutes, and then the RNA was purified by acid phenol extraction. The final pellet was resuspended in 35 µl of ddH₂O and frozen at -80 °C. The RNA quality was checked by electrophoresis on 1% agarose gels.

Estimation of nucleic acid concentration

An optical density of 1.0 at 260 nm is usually taken to be equivalent to a concentration of 50 µg/ml for double stranded DNA, 40 µg/ml for single stranded DNA and RNA, and approximately 20 µg/ml for single-stranded oligonucleotides samples. The ratio of values for optical densities measured at 260 nm and 280 nm is considered as 1.8 for pure sample of DNA and 2 for RNA and these are reduced by protein contaminants (Sambrook et al., 1989). Therefore, these values were used to determinate not only the concentration but also the purity of the samples.

Enzymatic modification of DNA

Restriction enzymes

Restriction endonucleases were used in the construction and analysis of recombinant plasmids. Each restriction enzyme functions optimally in a buffer of specific ionic strength. All buffers were supplied by the same company that supplied the enzymes and were used according to the manufacturer's instructions.

For analytical digests 100-500 ng of DNA were digested in a volume of 20 µl containing 6 U of the appropriate restriction enzyme. The reaction was incubated for 3-6 hours at 37 °C. Preparative digestion was made of 5-20 µg DNA using the above conditions and 2 U of restriction enzyme for µg of DNA in 200 µl reaction volume.

Large fragment of E. coli Polymerase I and T4 Polynucleotide Kinase

These enzymes were used to treat PCR products for blunt-end ligation during construction of recombinant plasmids. The large fragment of DNA Polymerase I (Klenow) is a proteolytic product of *E. coli* DNA Polymerase I. It retains polymerization and 3'->5' exonuclease activity, but has lost 5'->3' exonuclease activity. This was useful for digesting a specific A residues added by Taq DNA polymerase at the 3' terminus to create compatible ends for ligation. T4 Polynucleotide Kinase catalyses the transfer of phosphate from ATP to the 5' hydroxyl terminus of DNA. It was useful for the addition of 5'-phosphate to PCR products to allow subsequent ligation. Klenow fragment (2.5 U) was added to 23 µl of PCR product in 5 mM MgCl₂ buffer. The mixture was incubated at room temperature for 10 minutes. EDTA to a final concentration of 0.2 mM, ATP to a final concentration of 1 mM, 10 U of T4 Polynucleotide Kinase and the proper quantity of Kinase buffer were added to the above

mixture and incubated at 37 °C for 30 min. The enzymes were inactivated by incubation at 80 °C for 20 min.

T4 DNA ligase

T4 DNA ligase catalyses the formation of a phosphodiester bond between adjacent 3' hydroxyl and 5' phosphoryl termini in DNA, requiring ATP as a cofactor in this reaction. This enzyme was used to join double stranded DNA fragments with compatible sticky or blunt ends, during generation of recombinant plasmid DNAs.

20 ng of linearized vector were ligated with a 5-10 fold molar excess of insert in a total volume of 30 µl containing 1X ligase buffer and 1U of T4 DNA ligase. Reaction was carried out at room temperature for 6-12 hours.

In some reactions synthetic oligonucleotide were included in the reaction. In these cases, the amounts added to each reaction to obtain inclusion of oligonucleotides in the resulting plasmid were about 100 fold molar excess over the DNA vector.

Agarose gel electrophoresis of DNA

DNA samples were size fractionated by electrophoresis in agarose gels ranging in concentrations from 0.8 % w/v (large fragments) to 2 % w/v (small fragments). The gels contained ethidium bromide (0.5 µg /ml) and 1X TBE. Horizontal gels were routinely used for fast analysis of DNA restriction enzyme digests, estimation of DNA concentration, or DNA fragment separation prior to elution from the gel. Samples of 20 µl containing 1X DNA loading buffer were loaded into submerged wells. The gels were electrophoresed at 50-80 mA in 1X TBE running buffer for a time depending on the fragment length expected and gel concentration. DNA was visualized by UV transillumination and the result recorded by digital photography.

Elution and purification of DNA fragments from agarose gels

This protocol was used to purify small amounts (less than 1 µg) of DNA for sub-cloning. The DNA samples were electrophoresed onto an agarose gel as described previously. The DNA was visualized with UV light and the required DNA fragment band was excised from the gel. This slab was cut into pieces, and the JETquick Spin Column Technique (Genomed) was used according to the manufacturer's instructions. Briefly, 300 µl of gel solubilisation solution L1 (NaClO₄, Na acetate and TBE) were added for each 100 mg of the gel slice pieces and incubated at 65 °C for 15 min vortexing every 5 min. The mixture was loaded into a prepared JETquick column and it was centrifuged at maximum speed for 2 min. The flowthrough was discarded. 700 µl of washing and reconstituted solution L2 (ethanol, NaCl, EDTA and Tris-HCl) were added into the spin column and after 5 min, the column was centrifuged in the same conditions twice. The flowthrough was again discarded both times. To elute the bound DNA, 30-50 µl of sterile water were added onto the centre of the silica matrix of the spin column and the system was centrifuged for 2 min. The amount of DNA recovered was approximately calculated by UV fluorescence of intercalated ethidium bromide in an agarose gel electrophoresis.

Preparation of bacterial competent cells

Bacterial competent cells were prepared following the method described by Chung and Niemela (Chung et al., 1989). *E. Coli* strains were grown overnight in 3 ml of LB at 37°C. The following day, 300 ml of fresh LB were added and the cells were grown at room temperature for 4-5 h until the OD₆₀₀ was 0.3-0.4. The cells were then put in ice and centrifuged at 4 °C and 1000g for 15 min. The pellet was resuspended in 30 ml of cold TSS solution (10% w/v PEG, 5% v/v DMSO, 35mM Mg Cl₂, pH 6.5 in LB medium). The cells were aliquoted, rapidly

freeze in liquid nitrogen and stored at -80°C . Competence was determined by transformation with 0.1 ng of pUC19 and was deemed satisfactory if this procedure resulted in more than 100 colonies.

Transformation of bacteria

Transformations of ligation reactions were performed using 1/2 of the reaction volume. Transformation of clones was carried out using 1 ng of the plasmid DNA. The DNA was incubated with 60 μl of competent cells for 20 min on ice and at 42°C for 2 minutes. At this point 60 μl of LB were added and the bacteria allowed to recover for 10 min at 37°C . The cells were then spread onto agarose plates containing the appropriate antibiotic. The plates were then incubated for 12-15 hours.

When DNA inserts were cloned into β -galactosidase-based virgin plasmids, 30 μl of IPTG 100 mM and 20 μl of X-Gal (4 % w/v in dimethylformamide) were spread onto the surface of the agarose before plating to facilitate screening of positive clones (white colonies) through identification of β -galactosidase activity (blue colonies).

Amplification of selected DNA fragments

The polymerase chain reaction was performed on genomic or plasmid DNA following the basic protocols of the Roche Diagnostic Taq DNA Polymerases. The volume of the reaction was 50 μl . The reaction buffer was: 1X Taq buffer, dNTP mix 200 μM each, oligonucleotide primers 100 nM each, Taq DNA Polymerase 2.5 U. As DNA template, 0.1 ng of plasmid or 100-500 ng of genomic DNA were used for amplification. When a DNA fragment longer than 2000bp was amplified, DMSO 3% was also added to the mixture. The amplification conditions are described for each particular PCR. The amplifications were performed on a

Cetus DNA Thermal Cycler (Perkin Elmer) or on a Gene Amp PCR System (Applied Biosystems).

Sequence analysis for cloning purpose

Sequence analysis of plasmid DNA was performed using the CEQ 2000 sequencer (Beckman Coulter). The plasmid DNA of interest (approximately 100 ng) was purified through a MicroSpin S-400 HR Column (Amersham Pharmacia Biotech). The DNA was then amplified using fluorescent labeled dideoxy nucleotide terminators according to the manufacturer's instructions. The samples were analyzed by loading them into the automatic sequencer.

Reconstruction of the whole Apo AII gene sequence

In order to study the mechanism of the Apo AII exon 3 splicing the full sequence of the gene was cloned XhoI-SacII in pBluescript. Each fragment of the complete Apo AII gene was amplified by PCR (94 °C 30 sec, 56 °C 1 min, 72 °C 1 min, 35 cycles) using approximately 200-400 ng of genomic DNA as template. After amplification, the PCR products were purified through a MicroSpin S-400 HR Column (Amersham Pharmacia Biotech, Sweden). The DNA was then directly sequenced by CEQ 2000 sequencer (Beckman) according to the manufacturer's instructions as described above for sequence analysis of plasmid DNA. The oligonucleotides listed below were used to amplify the whole Apo AII sequence from genomic DNA.

Name of the oligo	Sequence of the oligo (5'-3')
hApoAII-1173 XhoI 5'	tccgctcgagcccgggaggtggaggttgca
hApoAII intron2/Sal I 3'	tttggttgccagtcgacctggatataaga
hApoAII intron2/Sal I 5'	tcttataatccaggtcgactggcaacaaaa
hApoAII intron3/EcoR I 3'	tttctctggaattctaagcctaattct
hApoAII intron3/EcoR I 5'	agattaggcttagaattcgagagaaa
hApoAII stop as+Cla/Not	aatggtctggacactgcggccgcattcactgggtggcagggc
hApoAII stop 5'+Cla/Not	agcctgccaccagtgatgatgcggccgcagtgccagaccatt
hApoAII polyA SacII 3'	ttccccgaggtaggagactctgggtttgga

Each fragment was then ligated to the next up to totally reconstruct the gene (named as pApo-wt). Suitable SalI, EcoRI and NotI-ClaI restriction sites were included in intron 2, intron 3 and exon 4 respectively. The first two to create an interchangeable cassette to test the mutations or replacing made in the exon 3 and/or its flanking introns, whereas NotI-ClaI in exon 4 to differentiate pApo-wt from the endogenous Apo AII.

Oligos sequence of the different construct

Generation of CF/Apo hybrids

The oligos used to create the CF/Apo hybrids were:

Name of the oligo	Sequence of the oligo (5'-3')
Apo/CF short Sense	aacagggatttgggaattatttaggcaaaggagccatgtgtgg
Apo/CF short AS	ccacacatggctccttgcctaaataattccccaatccctgtt
Apo/CF long Sense	ctcacttcttggtactcctgtcctgagcttcaggccgaggccaagt
Apo/CF long AS	acttggcctcggcctgaagctcaggacaggagtaccaagaagtgg
hCF 9 +60 / hAII Sense	atagaaaaacttctaattccagaccgtgactgac
hCF 9 +60 / hAII AS	gtcagtcacggtctggaattagaagttttctat
HAPO-CF SHORT/MEDIUM S	agcaaaacaaaacaatagagcctggttct
HAPO-CF SHORT/MEDIUM AS	agaaaccaggctctattgtttgtttgct
HAPO-CF MEDIUM/LONG S	ctcttctcagtaatacctgatggagaagg
HAPO-CF MEDIUM/LONG AS	ccttctccatcaggtattactgaagaagag

Generation of overlapping deletion within Apo AII exon 3

The oligos used to create the overlapping deletions were:

Name of the oligo	Sequence of the oligo (5'-3')
ApoAII-Ex3 D1840-61 S	accgtgactgactatgtcaagagcccagag
ApoAII-Ex3 D1840-61 AS	ctctgggctcttgacatagtcagtcacggt
ApoAII-Ex3 D1854-75 S	tatggcaaggacctggagcttcaggccgag
ApoAII-Ex3 D1854-75 AS	ctcggcctgaagctccaggtccttgccata
ApoAII-Ex3 D1866-86 S	ctgatggagaaggctcaggccaagtaagtc
ApoAII-Ex3 D1866-86 AS	gacttacttggcctcgaccttctccatcag

Generation of point mutation within Apo AII exon 3 ESE

The oligos used to point mutate the Apo AII exon 3 ESE were:

Name of the oligo	Sequence of the oligo (5'-3')
91-T->A S	tggcaaggaccagatggagaaggtaaga
91-T->A AS	tcttgaccttctccatctgggccttgcca
92-G->C S	tggcaaggaccacatggagaaggtaaga
92-G->C AS	tcttgaccttctccatctgggccttgcca
93-A->T S	tggcaaggaccagttggagaaggtaaga
93-A->T AS	tcttgaccttctccaactgggccttgcca
94-T->A S	tggcaaggaccagaaggagaaggtaaga
94-T->A AS	tcttgaccttctccttctgggccttgcca
95-G->C S	tggcaaggaccagatcgagaaggtaaga
95-G->C AS	tcttgaccttctcgatctgggccttgcca
96-G->C S	tggcaaggaccagatgcagaaggtaaga
96-G->C AS	tcttgaccttctgcactctgggccttgcca
97-A->T S	tggcaaggaccagatgggaaggtaaga
97-T->A AS	tcttgaccttcaccatctgggccttgcca
98-G->C S	tggcaaggaccagatggacaaggtaaga
98-G->C AS	tcttgaccttgtccatctgggccttgcca
99-A->T S	tggcaaggaccagatggagtaggtaaga
99-A->T AS	tcttgaccttctccatctgggccttgcca

Generation of mutagenesis of G runs in intron 3

The oligos used to mutagenize the Gs runs in Apo AII intron 3 were:

Name of the oligo	Sequence of the oligo (5'-3')
Apo m IVS3 Gn S	gtaagtctcatcacaagcgcttca-
Apo m IVS3 Gn AS	tgaagcgcttggatgagacttac
Apo IVS3 mut G1-3 S	gtaagtctcatgtcaagcattcatgcgtgtg
Apo IVS3 mut G1-3 AS	cacagcgcatgaatgcttgacatgagacttac
ApoAII-Delta G4CTG 5'	ggcaaggggttcagagcttatggaactgtggaga
ApoAII-Delta G4CTG 3'	tctccacagttccataagctctgaaccccttgcc

Generation of Apo AII exon 3 5'splice site mutants

The oligos used to mutagenize the Apo AII exon 3 5'splice site were:

Name of the oligo	Sequence of the oligo (5'-3')
ApoAII-Ex3 D5' S	caggccgaggccgacatactctcagggcaa
ApoAII-Ex3 D5' AS	ttgccctgagagtatgtcggcctcggcctg

Replacement of EDA pY tract by GT repeats

Wild type pY tract of the EDA exon in the α -globin/fibronectin reporter system (Muro et al., 1998) was replaced by (GT)₁₆ repeats. Sense and antisense oligos carrying a (GT)₁₆ were designed such that annealed at the flanking sequences of the EDA pY tract. A first PCR reaction (94°C-30sec, 60°C-30sec, 72°C-30sec for 30 cycles) was performed by using a sense oligo annealing at -1 EDA exon and the antisense oligo carrying the (GT)₁₆ sequence. A second reaction (same PCR conditions) was carried out by using the sense oligo carrying the (GT)₁₆ sequence and the antisense oligo annealing at +1 EDA exon. Afterward a third PCR reaction was carried out (same conditions) by using the primers annealing at the extremes Ex-1S and Ex+1AS and as a template, 1 μ L of reaction 1 and reaction 2. The final "extended product" was visualized in a 1.0% agarose gel, purified, cloned and the plasmid transfected in NIH 3T3 cells. The primers utilized to generate this construct are listed below:

Reaction n°	Name of the oligo	Sequence of the oligo (5'-3')
1	Ex-1S	atcaaacagaaatgacc attga
1	FN-EDArev	atcaatgtctgttaggcaaattaatggtaagagggttatgtgaaaagcaagttgtg
1	FN-AII rev	atcaatgtctgcacacacacacacacacacacacacacacactctgttgtg
2	FN-EDAdir	cacaacttgcttttcacataacctcttaccattaattgcctaacagacattgat
2	FN-AII dir	cacaacagagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgcagacattgat
2	Ex+1AS	ctctttcgggttcacccgca

Intron removal order experiment

In order to determine the order in which ApoAII introns are processed during the splicing reaction a similar strategy used by Kessler et al (Kessler et al., 1993). was carried out. PCR amplification (94°C-30sec, 60°C-1.0 min, 72°C-1.5 min for 35 cycles) of a Hep3B non-transfected cells cDNA, was performed by using the oligonucleotides listed below annealing in the exon 1 / intron 3 (reaction1) and intron 2/ exon 4 (reaction 2).

Reaction n°	Name of the oligo	Sequence of the oligo (5'-3')
1	Ex 1-1221 S	accaaggacagagacgctggct
1	ApoAII-IVS3_1910-36AS	ccacagttccacagccctga
2	ApoAII-IVS2_1565-75S	agggctcaaaaggaagatca
2	Rev Cla/Not	ttttgcggccgcacgcgattcactcggtggcaggctgtgttccaa

Oligo antisense for TDP-43 (TIO1318)

The antisense phosphorotioate (PS)-oligonucleotides sequence was:

Name of the oligo	Sequence of the oligo (5'-3')
TIO1318	ctgtctacattccccagcca

Oligos for Electro Mobility Shift Assay (EMSA)

Name of the oligo	Sequence of the oligo (5'-3')
ESE wt S	ggtacctatggcaaggacctgatggagaaggtaagagaagctt
ESE wt AS	aagcttctcttgaccttctccatcaggtccttgccataggtacc
ΔESE S	ggtacctatggcaaggaccaggtcaagagaagctt
ΔESE AS	aagcttctcttgacctggctccttgccataggtacc

Oligos for UV crosslinking

Name of the oligo	Sequence of the oligo (5'-3')
ESE wt S	ggtacctatggcaaggacctgatggagaaggtaagagaagctt
ESE wt AS	aagcttctcttgaccttctccatcaggtccttgccataggtacc
ΔESE S	ggtacctatggcaaggaccaggtcaagagaagctt
ΔESE AS	aagcttctcttgacctggccttgccataggtacc
ESE A97T S	ggtacctatggcaaggacctgatggtgaaggtaagagaagctt
ESE A97T AS	aagcttctcttgaccttcaccatcaggtccttgccataggtacc

Oligos for UV crosslinking (ISE within intron 3)

Name of the oligo	Sequence of the oligo (5'-3')
IVS3-ISE-S	gcgtaatacgactcactatagggagactgtggaactgtggagagaaagaagggaagatgagaggt cccacagaagtctgaaccca
IVS3-ISE-AS	tgggttcagacttctgtgggacctctcatcttccttcttctctccacagttcc acagtctccctatagtgagtcgtattacgc
IVS3-ISE-Mut-S	gcgtaatacgactcactatagggagactgtggaactgtcgtgacaatgacggtaacatta gcgggtcccacagaagtctgaaccca
IVS3-ISE-Mut-AS	tgggttcagacttctgtgggaccgctaattgtaccgtcattgtcacgacag ttccacagtctccctatagtgagtcgtattacgc

Construction of the minigene system

Apo AII exon 2-intron 2-exon 3 minigene used for UV-crosslinking

The 629bp human Apo AII sequence encompassing part of exon 2 (base 1401) to intron 3 (base 2029) was amplified from pApo-wt by PCR (94 °C 30 sec, 60 °C 30 sec, 72 °C 30 sec, 30 cycles by using the oligonucleotides listed below

Name of the oligo	Sequence of the oligo (5'-3')
S XhoI	cccgctcgagcactgttaccacatgaagct
HindIII/SacII	tttccgcggaagcttctctgcgaattctaagcctaattcgcctaa

The first and second oligo include *XhoI* and *HindIII-SacII* target site respectively. PCR product was digested *XhoI-SacII* and cloned in pBluescript. Sequencing excluded the presence of mutations within the insert. Plasmid was then *HindIII* linearized for in vitro transcription with T7 RNAPolymerase.

Maintenance and analysis of cells in culture

Hep3B, NIH3t3 and HeLa cell lines were grown in Dulbecco's Mem with Glutamax I (Gibco) (Dulbecco's modified Eagle's medium with glutamine, sodium pyruvate, pyridoxine and 4.5 g/l glucose) supplemented with 10% foetal calf serum (Euro Clone) and Antibiotic Antimycotic (Sigma) according to the manufacturer's instructions.

Plates containing a confluent monolayer of cells were treated with 0.1% w/v trypsin as follows. Cells washed with PBS solution, were incubated at 37 °C with 1-2 ml of PBS/EDTA/trypsin solution (PBS containing 0.04% w/v EDTA and 0.1% w/v trypsin) for 2 minutes or until cells were dislodged. After adding 10 ml of media, cells were pelleted by centrifugation and resuspended in 5 ml pre-warmed medium. 1-2 ml of this cell suspension was added to 10 ml medium in a fresh plate and was gently mixed before incubation.

Transfections

The DNA used for transfections was prepared with JetStar purification kit (Genomed) as previously described. Liposome-mediated transfections of 3×10^5 human hepatocarcinoma Hep3B cells were performed using DOTAP Liposomal Transfection Reagent (Alexis Biochemicals). 3 μ g of construct DNA was mixed with 5 μ g of DOTAP for each transfection and the mixture was incubated at room temperature for 15 minutes to allow the formation of DNA-liposome complexes. The mixture was added to the cells in 3 ml of serum free culture medium and incubated at 37 °C. After 12 h the medium was replaced with fresh medium and 24 h later, the cells were harvested. RNA isolation followed as described.

Cotransfections

In cotransfection experiments, cells were transfected with 4 μ g of pApo-wt or pApo-A97T and the PS-oligo TIO1318 at a final concentration of 7,2 μ M or with 3 μ g of the TDP-43 expressing plasmid. For the SR proteins and hnRNPA1 overexpression experiments, construct Δ 1866-86 (2 μ g) was cotransfected with 1 and 2 μ g of SRp 20, SRp55 and SRp75. Construct pApo-wt (2 μ g) was cotransfected with 1 and 2 μ g of SR proteins ASF/SF2, SC35 and hnRNPA1.

mRNA Analysis by Polymerase Chain Reaction

cDNA synthesis

In order to synthesize cDNA, the 3 μ g of total RNA extracted from cells were mixed with Not I-d(T)₁₈ primer (Pharmacia) in a final volume of 20 μ l. After denaturation at 90°C the RNA and the primer were incubated for 1 hour at 37 °C in the following solution: 1X First Strand Buffer (Gibco), 10 mM DTT, 1 mM dNTPs, RNase inhibitor 20 U (Ambion) and Moloney murine leukemia virus reverse transcriptase 100 U (Gibco). 1 μ l of the cDNA reaction mix was used for the PCR analysis.

cDNA analysis

PCR analysis of cDNA was carried out for 35 cycles (94 °C 30 sec, 60 °C 1 min, 72 °C 1,5 min) in 50 µl reaction volumes. ApoAII full sequence and mutations/deletions in the different constructs were tested by these specific oligonucleotides:

Name of the oligo	Sequence of the oligo (5'-3')
Ex 1-1221 S	accaaggacagagacgctggct
Not/Cla rev	tctggacactcggccgcatcg

For amplification of the endogenous ApoAII, the oligos used were:

Name of the oligo	Sequence of the oligo (5'-3')
Ex 1-1221 S	accaaggacagagacgctggct
Rev Cla/Not	tttgcggccgcatcgattcactcggcaggctgtgttccaa

Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were added to protein sample buffer (2X final). Conventional slab gel SDS PAGE (Laemmli, 1970) was performed in vertical gels with the required percentage of polyacrylamide (37,5:1 acrylamide:bis-acrylamide, ProtoGel, National Diagnostics), depending on each case. The gels were run at 40 mA in 1X SDS-PAGE running buffer (50 mM Tris, 0.38 M glycine, 0.1 % w/v SDS). After running, gels were either stained with coomassie Blue R250 in methanol-water-acetic acid 45:45:10 (v/v/v).

RNA preparation

DNA template for in vitro transcription used for UV crosslinking experiments, EMSA and in vitro splicing experiments, were prepared with oligos detailed above.

The templates for EMSA and UV-crosslinking were cloned into Kpn I and Hind III digested pBluescript SK under the T7 promoter control in pBluescript, whereas the templates for in vitro splicing were cloned into XhoI-PvuII (for the polypyrimidine tract replacement by GT repeats) or XhoI-BamHI (when the Tropomyosin exon 3 besides its pY tract were replaced by the Apo exon 3 and the (GU)_n, respectively) digested PY7 under the SP6 promoter control.

Transcription with T7 or SP6 RNA polymerase (Promega) was performed using 1-2 µg of linearized plasmid, in the presence of α -³²P-UTP. The reaction consisted of 1X transcription buffer (Promega), 40 U of RNase inhibitor (Ambion), 1.5 mM each of the three NTPs (-UTP), and 10 U of T7 RNA polymerase (Promega) in a final volume of 20 µl. Following incubation for 1 h at 37 °C the labeled RNA was then purified on a Nick column (Pharmacia), precipitated and resuspended in RNase-free water. The specific activities were in the range of 4×10^6 c.p.m./mg of RNA.

RNA EMSA

The Electrophoretic Mobility Shift Assay (EMSA) method is described in Figure 4.1. In vitro T7 transcribed ^{32}P -labelled RNA (4-6 fmol) was incubated with 5.2 mM Hepes pH 7.9, 1 mM MgCl_2 , 0.8 mM magnesium acetate, 0.52 mM DTT, 3.8% glycerol, 0.75 mM ATP, 1 mM GTP, 0.5 $\mu\text{g}/\mu\text{l}$ of heparin and 30 μg HeLa Nuclear

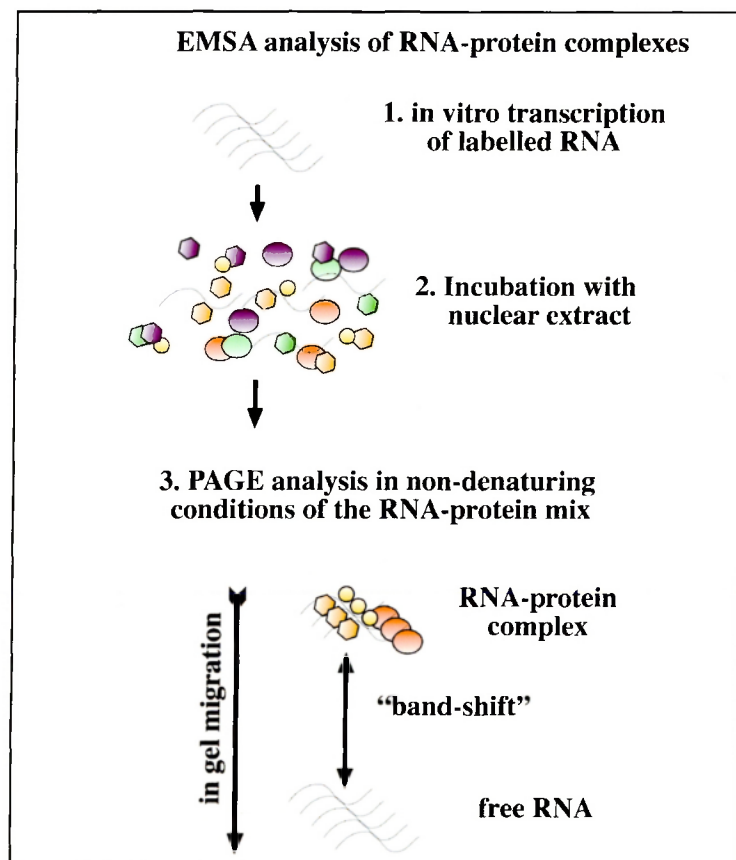


Figure 4.1 Electrophoretic Mobility Shift Assay (EMSA).

extract (4C Biotech) in a final volume of 20 μ l for 20 min at room temperature. Following the addition of 5 μ l of 50% (v/v) glycerol and tracking dye, complexes were resolved on a 4% polyacrylamide gel (ratio of 19:1 acrylamide:bis) in 75 mM Tris-Glycine buffer (75 mM Tris, 75 mM Glycine) at 15-25 mA for 3-4 h at 4 °C. Gels were dried and exposed to X-OMAT AR films for 1-3 h.

UV cross-linking

The UV cross-linking technique is described in Figure 4.2. The assays were carried out with nuclear extracts. Apo AII minigenes cloned in pBluescripts SK plasmids were linearized by Hind III digestion and labelled RNAs were transcribed as described previously. 32 P-labeled RNA (1×10^6 cpm per incubation) was incubated with 30 μ g HeLa Nuclear extract (4C Biotech) in 20 μ l final volume.

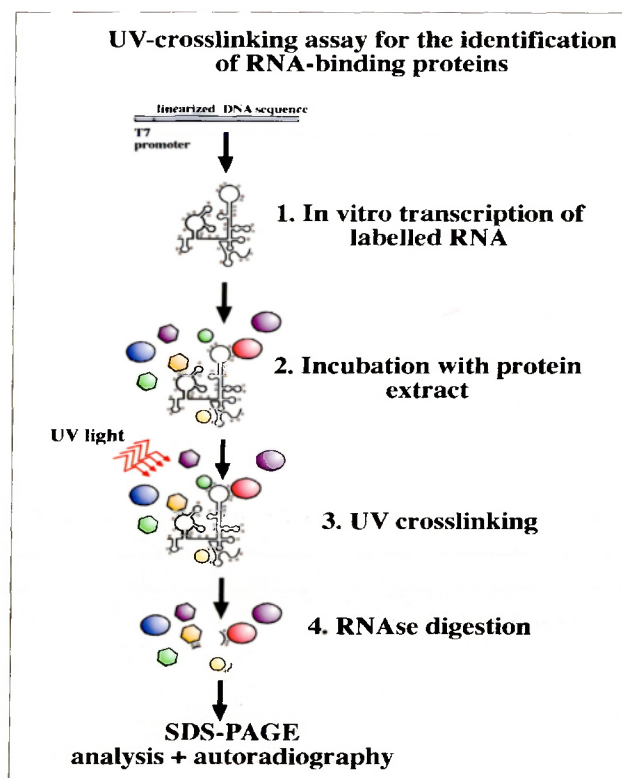


Figure 4.2 Scheme of UV cross-linking assay.

Final binding conditions were 5.2 mM Hepes pH 7.9, 1 mM MgCl₂, 0.8 mM magnesium acetate, 0.52 mM DTT, 3.8% glycerol, 0.75 mM ATP and 1 mM GTP. When UV cross-linking assays were performed with ISE-wt and ISE-mut constructs, also heparin at a 5 µg/µl final concentration was added to the binding mixture. The mixture was incubated for 15 min at 30 °C. Following incubation, the samples were irradiated with UV light on ice (0.8 joules, 5 min) using a UV Linker (Euroclone). Unbound RNA was removed by digestion with 1 U of RNase A at 37 °C for 30 min. The resulting ³²P-labeled proteins were resolved by 10% SDS-PAGE in the presence of molecular weight markers. The gel was dried and exposed to X-OMAT-AR film for 2-5 days.

Immunoprecipitation of SR proteins following UV-crosslinking.

The RNAs probes used in this study (ESE-wt, ESEA97T and ΔESE) were obtained by cloning in Bluescript SK+ the human ApoAII sequence ranging from intron 2 (nt 1549) to intron 3 (nt 2022) with the wild type ESE, mutation A97T ESE and ΔESE. Construct carrying the EDA exon (from nucleotide 107 to 270) was used as a positive control. The UV cross-linking/immunoprecipitation assay is described in Figure 4.3. Each plasmid was then linearized with the appropriate restriction enzyme and transcribed using T7 RNA polymerase according to standard conditions. The UV crosslinking of [α³²P]-UTP labelled RNAs with commercial HeLa nuclear extract (C4, Biotech) were performed as describe before. To each sample we added 150 µL of IP buffer (20 mM Tris pH=8.0, 300mM NaCl, 1mM EDTA, 0.25% NP-40) together with 1 µg of monoclonal antibodies and incubated for 2 hours at 4°C on a rotator wheel. Anti-SF2/ASF (mAb96) and Anti-SC35 monoclonal antibodies were purchased from Zymed Laboratories Inc and Sigma respectively. After 2 hours incubation we added to each sample 30 µL of protein A7G-Plus Agarose (Santa Cruz Biotechnologies) and the incubation was left at 4°C overnight. The following morning the beads were subjected to

four washing cycles with 1.5mL of IP buffer and were then loaded on a 10% SDS-PAGE gel. Gels were run at a constant 30mA for approx. 3 hours and a half, were dried and then exposed for 4-6 days with a Biomax Screen (Kodak).

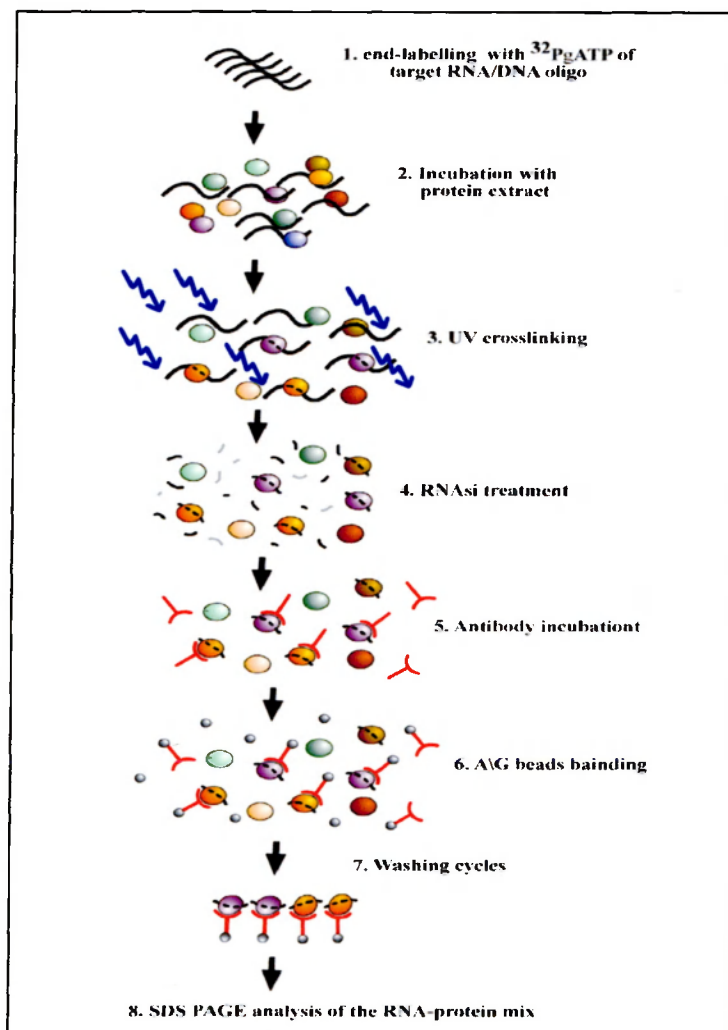


Figure 4.3 UV cross-linking/Immunoprecipitation assay

In vitro splicing assays

HeLa cell nuclear was purchased from C4, Biotech. Standard conditions were used for the splicing reactions (Mayeda and Krainer, 1999). 20 fmol (usually 0.1-0.4 μL) of ^{32}P -labeled 7-CH3 GpppG-capped SP6 transcripts was incubated in 10- μL splicing reactions. Each reaction contained 1.0 μL of a 25X ATP/CP mixture, 1.0 μL of 80 mM MgCl_2 (at a final concentration of 3.2mM), 1.25 μL of 0,4 M HEPES-KOH, 30% HeLa nuclear extract, pH 7.3, 5 μL of 13% PVA (added last) and Milli-Q water to 25 μL total. After incubation at 30°C for 2–4 hr, the RNA was extracted and analyzed on 8% polyacrylamide denaturing gels, followed by autoradiography.

REFERENCES

- Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M., and Luhrmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *Embo J* 18, 5789-5802.
- Ariffin, H., Millar, D. S., Cooper, D. N., Chow, T., and Lin, H. P. (2003). Prenatal exclusion of severe factor VII deficiency. *J Pediatr Hematol Oncol* 25, 418-420.
- Bai, Y., Lee, D., Yu, T., and Chasin, L. A. (1999). Control of 3' splice site choice in vivo by ASF/SF2 and hnRNP A1. *Nucleic Acids Res* 27, 1126-1134.
- Balvay, L., Libri, D., and Fiszman, M. Y. (1993). Pre-mRNA secondary structure and the regulation of splicing. *Bioessays* 15, 165-169.
- Bear, J., Tan, W., Zolotukhin, A. S., Tabernero, C., Hudson, E. A., and Felber, B. K. (1999). Identification of novel import and export signals of human TAP, the protein that binds to the constitutive transport element of the type D retrovirus mRNAs. *Mol Cell Biol* 19, 6306-6317.
- Bell, M. V., Cowper, A. E., Lefranc, M. P., Bell, J. I., and Screaton, G. R. (1998). Influence of intron length on alternative splicing of CD44. *Mol Cell Biol* 18, 5930-5941.
- Bentley, D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol* 11, 347-351.
- Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol* 14, 336-342.
- Berget, S. M. (1995). Exon recognition in vertebrate splicing. *J Biol Chem* 270, 2411-2414.
- Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 74, 3171-3175.

- Black, D. L. (1991). Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? *Genes Dev* 5, 389-402.
- Blencowe, B. J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 25, 106-110.
- Blencowe, B. J., Issner, R., Nickerson, J. A., and Sharp, P. A. (1998). A coactivator of pre-mRNA splicing. *Genes Dev* 12, 996-1009.
- Bouck, J., Fu, X. D., Skalka, A. M., and Katz, R. A. (1995). Genetic selection for balanced retroviral splicing: novel regulation involving the second step can be mediated by transitions in the polypyrimidine tract. *Mol Cell Biol* 15, 2663-2671.
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc Natl Acad Sci U S A* 75, 4853-4857.
- Brosi, R., Hauri, H. P., and Kramer, A. (1993). Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem* 268, 17640-17646.
- Brousseau, T., Dupuy-Gorce, A. M., Evans, A., Arveiler, D., Ruidavets, J. B., Haas, B., Cambou, J. P., Luc, G., Ducimetiere, P., Amouyel, P., and Helbecque, N. (2002). Significant impact of the highly informative (CA)_n repeat polymorphism of the APOA-II gene on the plasma APOA-II concentrations and HDL subfractions: The ECTIM study. *Am J Med Genet* 110, 19-24.
- Buratti, E., and Baralle, F. E. (2001). Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J Biol Chem* 276, 36337-36343.
- Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M., and Baralle, F. E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *Embo J* 20, 1774-1784.

- Burge, C., and Sharp, P. (1999). Splicing of precursor to mRNAs by the spliceosome. In *The RNA World*, C. S. H. L. Press, ed. (New York, edn. 2. C. T. Gesteland RF, Atkins JF.).
- Buvoli, M., Mayer, S. A., and Patton, J. G. (1997). Functional crosstalk between exon enhancers, polypyrimidine tracts and branchpoint sequences. *Embo J* 16, 7174-7183.
- Caceres, J. F., Stamm, S., Helfman, D. M., and Krainer, A. R. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* 265, 1706-1709.
- Caputi, M., Casari, G., Guenzi, S., Tagliabue, R., Sidoli, A., Melo, C. A., and Baralle, F. E. (1994). A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. *Nucleic Acids Res* 22, 1018-1022.
- Carlo, T., Sterner, D. A., and Berget, S. M. (1996). An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. *Rna* 2, 342-353.
- Cartegni L, C. S., Krainer A (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3, 285-298.
- Cartegni, L., and Krainer, A. R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat Genet* 30, 377-384.
- Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31, 3568-3571.
- Chambaz, J., Cardot, P., Pastier, D., Zannis, V. I., and Cladaras, C. (1991). Promoter elements and factors required for hepatic transcription of the human ApoA-II gene. *J Biol Chem* 266, 11676-11685.
- Champion-Arnaud, P., and Reed, R. (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev* 8, 1974-1983.

- Chandler, S. D., Mayeda, A., Yeakley, J. M., Krainer, A. R., and Fu, X. D. (1997). RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proc Natl Acad Sci U S A* *94*, 3596-3601.
- Chen, I. T., and Chasin, L. A. (1994). Large exon size does not limit splicing in vivo. *Mol Cell Biol* *14*, 2140-2146.
- Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* *12*, 1-8.
- Chung, T. C., Niemela, S. L., and Miller, R. H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A* *86*, 2172-2175.
- Civeira, F., Genest, J., Pocovi, M., Salem, D. N., Herbert, P. N., Wilson, P. W., Schaefer, E. J., and Ordovas, J. M. (1992). The *MspI* restriction fragment length polymorphism 3' to the apolipoprotein A-II gene: relationships with lipids, apolipoproteins, and premature coronary artery disease. *Atherosclerosis* *92*, 165-176.
- Coleman, T. P., and Roesser, J. R. (1998). RNA secondary structure: an important cis-element in rat calcitonin/CGRP pre-messenger RNA splicing. *Biochemistry* *37*, 15941-15950.
- Collins, C. A., and Guthrie, C. (1999). Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev* *13*, 1970-1982.
- Colwill, K., Feng, L. L., Yeakley, J. M., Gish, G. D., Caceres, J. F., Pawson, T., and Fu, X. D. (1996a). SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *J Biol Chem* *271*, 24569-24575.
- Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., and Duncan, P. I. (1996b). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *Embo J* *15*, 265-275.

- Cook, P. R. (1999). The organization of replication and transcription. *Science* 284, 1790-1795.
- Coolidge, C. J., Seely, R. J., and Patton, J. G. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res* 25, 888-896.
- Cooper DN, K. M., Antonarakis SE. (1995). The nature and mechanisms of human gene mutation. In *The Metabolic and Molecular Bases of Inherited Disease.*, B. A. Scriver CR, Sly WS, Valle D, eds., ed. (New York, NY, McGraw-Hill), pp. 259-291.
- Corden, J. L., and Patturajan, M. (1997). A CTD function linking transcription to splicing. *Trends Biochem Sci* 22, 413-416.
- Coulter, L., MA, L., and Cooper, T. (1997). Identification of a New class of Exonic Splicing Enhancers by In Vivo Selection. *Mol Cell Biol* 17, 2143-2150.
- Croft, L., Schandorff, S., Clark, F., Burrage, K., Arctander, P., and Mattick, J. S. (2000). ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nat Genet* 24, 340-341.
- Cuppens, H., Lin, W., Jaspers, M., Costes, B., Teng, H., Vankeerberghen, A., Jorissen, M., Droogmans, G., Reynaert, I., Goossens, M., *et al.* (1998). Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (Tg)m locus explains the partial penetrance of the T5 polymorphism as a disease mutation. *J Clin Invest* 101, 487-496.
- Damgaard, C. K., Tange, T. O., and Kjems, J. (2002). hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *Rna* 8, 1401-1415.
- Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell* 5, 779-787.

- Deeb, S. S., Takata, K., Peng, R. L., Kajiyama, G., and Albers, J. J. (1990). A splice-junction mutation responsible for familial apolipoprotein A-II deficiency. *Am J Hum Genet* 46, 822-827.
- Deirdre, A., Scadden, J., and Smith, C. W. (1995). Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *Embo J* 14, 3236-3246.
- Del Gatto, F., and Breathnach, R. (1995). Exon and intron sequences, respectively, repress and activate splicing of a fibroblast growth factor receptor 2 alternative exon. *Mol Cell Biol* 15, 4825-4834.
- Dirksen, W. P., Hampson, R. K., Sun, Q., and Rottman, F. M. (1994). A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. *J Biol Chem* 269, 6431-6436.
- Dominski, Z., and Kole, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* 11, 6075-6083.
- Dominski, Z., and Kole, R. (1992). Cooperation of pre-mRNA sequence elements in splice site selection. *Mol Cell Biol* 12, 2108-2114.
- Dominski, Z., and Kole, R. (1994). Identification and characterization by antisense oligonucleotides of exon and intron sequences required for splicing. *Mol Cell Biol* 14, 7445-7454.
- Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62, 289-321.
- Du, L., and Warren, S. L. (1997). A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. *J Cell Biol* 136, 5-18.
- Dupuy-Gorce, A. M., Desmarais, E., Vigneron, S., Buresi, C., Nicaud, V., Evans, A., Luc, G., Arveiler, D., Marques-Vidal, P., Cambien, F., *et al.* (1996). DNA polymorphisms in

- linkage disequilibrium at the 3' end of the human APO AII gene: relationships with lipids, apolipoproteins and coronary heart disease. *Clin Genet* 50, 191-198.
- Eick, D. (1990). Elongation and maturation of c-myc RNA is inhibited by differentiation inducing agents in HL60 cells. *Nucleic Acids Res* 18, 1199-1205.
- Estes, P. A., Cooke, N. E., and Liebhaber, S. A. (1992). A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms. *J Biol Chem* 267, 14902-14908.
- Fairbrother, W. G., and Chasin, L. A. (2000). Human genomic sequences that inhibit splicing. *Mol Cell Biol* 20, 6816-6825.
- Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. (2002). Predictive identification of exonic splicing enhancers in human genes. *Science* 297, 1007-1013.
- Fischer, U., Sumpter, V., Sekine, M., Satoh, T., and Luhrmann, R. (1993). Nucleocytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *Embo J* 12, 573-583.
- Fu, X. D. (1993). Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* 365, 82-85.
- Fu, X. D. (1995). The superfamily of arginine/serine-rich splicing factors. *Rna* 1, 663-680.
- Fu, X. Y., Colgan, J. D., and Manley, J. L. (1988). Multiple cis-acting sequence elements are required for efficient splicing of simian virus 40 small-t antigen pre-mRNA. *Mol Cell Biol* 8, 3582-3590.
- Galiana-Arnoux, D., Lejeune, F., Gesnel, M. C., Stevenin, J., Breathnach, R., and Del Gatto-Konczak, F. (2003). The CD44 alternative v9 exon contains a splicing enhancer responsive to the SR proteins 9G8, ASF/SF2, and SRp20. *J Biol Chem* 278, 32943-32953.

- Gallego, M. E., Gattoni, R., Stevenin, J., Marie, J., and Expert-Bezancon, A. (1997). The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A. *Embo J* 16, 1772-1784.
- Gattoni, R., Schmitt, P., and Stevenin, J. (1988). In vitro splicing of adenovirus E1A transcripts: characterization of novel reactions and of multiple branch points abnormally far from the 3' splice site. *Nucleic Acids Res* 16, 2389-2409.
- Gersappe, A., and Pintel, D. J. (1999). CA- and purine-rich elements form a novel bipartite exon enhancer which governs inclusion of the minute virus of mice NS2-specific exon in both singly and doubly spliced mRNAs. *Mol Cell Biol* 19, 364-375.
- Goguel, V., and Rosbash, M. (1993). Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* 72, 893-901.
- Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* 10, 233-243.
- Graveley, B. R., Hertel, K. J., and Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre- mRNA splicing enhancers. *Embo J* 17, 6747-6756.
- Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K., and Hutton, M. (1999). 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J Biol Chem* 274, 15134-15143.
- Gui, J. F., Tronchere, H., Chandler, S. D., and Fu, X. D. (1994). Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc Natl Acad Sci U S A* 91, 10824-10828.

- Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C., and Mattaj, L. W. (1994). The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* 76, 531-541.
- Guzik, B. W., Levesque, L., Prasad, S., Bor, Y. C., Black, B. E., Paschal, B. M., Rekosh, D., and Hammarskjold, M. L. (2001). NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of intron-containing RNA in mammalian cells. *Mol Cell Biol* 21, 2545-2554.
- Hampson, R. K., La Follette, L., and Rottman, F. M. (1989). Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences. *Mol Cell Biol* 9, 1604-1610.
- Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., Jr., and Krainer, A. R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *Rna* 4, 430-444.
- Hastings, M. L., and Krainer, A. R. (2001). Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *Rna* 7, 471-482.
- Hastings, M. L., Wilson, C. M., and Munroe, S. H. (2001). A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. *Rna* 7, 859-874.
- Hawkins, J. D. (1988). A survey of intron and exon length. *Nucleic Acid Research* 16, 9893-9908.
- Hellman, L., Steen, M. L., Sundvall, M., and Pettersson, U. (1988). A rapidly evolving region in the immunoglobulin heavy chain loci of rat and mouse: postulated role of (dC-dA)_n.(dG-dT)_n sequences. *Gene* 68, 93-100.
- Hoffman, B. E., and Grabowski, P. J. (1992). U1 snRNP targets an essential splicing factor, U2AF₆₅, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev* 6, 2554-2568.

- Howe, K. J. (2002). RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys Acta* 1577, 308-324.
- Huh, G. S., and Hynes, R. O. (1993). Elements regulating an alternatively spliced exon of the rat fibronectin gene. *Mol Cell Biol* 13, 5301-5314.
- Hussain, M. M., Zannis, V. I. (1990). Intracellular modification of human apolipoprotein (apoA-II) and sites of apoA-II synthesis: comparison of apoA-II with apoC-II and apoC-III isoproteins. *Biochemistry* 29, 209-217.
- Ikeda, Y., Takagi, A., Nakata, Y., Sera, Y., Hyoudou, S., Hamamoto, K., Nishi, Y., and Yamamoto, A. (2001). Novel compound heterozygous mutations for lipoprotein lipase deficiency. A G-to-T transversion at the first position of exon 5 causing G154V missense mutation and a 5' splice site mutation of intron 8. *J Lipid Res* 42, 1072-1081.
- Izaurrealde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* 78, 657-668.
- Izaurrealde, E., and Mattaj, I. W. (1995). RNA export. *Cell* 81, 153-159.
- Kambach, C., Walke, S., and Nagai, K. (1999). Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol* 9, 222-230.
- Kamma, H., Portman, D. S., and Dreyfuss, G. (1995). Cell type-specific expression of hnRNP proteins. *Exp Cell Res* 221, 187-196.
- Kanopka, A., Muhlemann, O., and Akusjarvi, G. (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre- mRNA. *Nature* 381, 535-538.
- Katahira, J., Strasser, K., Podtelejnikov, A., Mann, M., Jung, J. U., and Hurt, E. (1999). The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *Embo J* 18, 2593-2609.

- Kataoka, N., Yong, J., Kim, V. N., Velazquez, F., Perkinson, R. A., Wang, F., and Dreyfuss, G. (2000). Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* 6, 673-682.
- Kazazian, H. H., Jr. (1998). Mobile elements and disease. *Curr Opin Genet Dev* 8, 343-350.
- Kazazian, H. H., Jr., and Moran, J. V. (1998). The impact of L1 retrotransposons on the human genome. *Nat Genet* 19, 19-24.
- Kessler, O., Jiang, Y., and Chasin, L. A. (1993). Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol Cell Biol* 13, 6211-6222.
- Kim, V. N., and Dreyfus, G. (2001). Nuclear mRNA binding proteins couple pre-mRNA splicing and post- splicing events. *Mol Cells* 12, 1-10.
- Kim, V. N., Kataoka, N., and Dreyfuss, G. (2001). Role of the nonsense-mediated decay factor hUpf3 in the splicing- dependent exon-exon junction complex. *Science* 293, 1832-1836.
- Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., and Manley, J. L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368, 119-124.
- Konarska, M. M., Grabowski, P. J., Padgett, R. A., and Sharp, P. A. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* 313, 552-557.
- Konarska, M. M., and Sharp, P. A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* 49, 763-774.
- Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* 65, 367-409.

- Krawczak, M., Reiss, J., and Cooper, D. N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* 90, 41-54.
- Kuo, H. C., Nasim, F. H., and Grabowski, P. J. (1991). Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* 251, 1045-1050.
- Kwok, J. B., Halliday, G. M., Brooks, W. S., Dolios, G., Laudon, H., Murayama, O., Hallupp, M., Badenhop, R. F., Vickers, J., Wang, R., *et al.* (2003). Presenilin-1 mutation L271V results in altered exon 8 splicing and Alzheimer's disease with non-cored plaques and no neuritic dystrophy. *J Biol Chem* 278, 6748-6754.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Lang, K. M., and Spritz, R. A. (1983). RNA splice site selection: evidence for a 5' leads to 3' scanning model. *Science* 220, 1351-1355.
- Langford, C. J., Klinz, F. J., Donath, C., and Gallwitz, D. (1984). Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell* 36, 645-653.
- Lavigne, A., La Branche, H., Kornblihtt, A. R., and Chabot, B. (1993). A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes Dev* 7, 2405-2417.
- Le Hir, H., Gatfield, D., Braun, I. C., Forler, D., and Izaurralde, E. (2001). The protein Mago provides a link between splicing and mRNA localization. *EMBO Rep* 2, 1119-1124.
- Le Hir, H., Izaurralde, E., Maquat, L. E., and Moore, M. J. (2000a). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *Embo J* 19, 6860-6869.

- Le Hir, H., Moore, M. J., and Maquat, L. E. (2000b). Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev* 14, 1098-1108.
- Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., and Mattaj, I. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev* 10, 1683-1698.
- Lim, L. P., and Burge, C. B. (2001). A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A* 98, 11193-11198.
- Lin, R. J., Newman, A. J., Cheng, S. C., and Abelson, J. (1985). Yeast mRNA splicing in vitro. *J Biol Chem* 260, 14780-14792.
- Litt, M., and Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44, 397-401.
- Liu, H. X., Cartegni, L., Zhang, M. Q., and Krainer, A. R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet* 27, 55-58.
- Liu, H. X., Chew, S. L., Cartegni, L., Zhang, M. Q., and Krainer, A. R. (2000). Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol* 20, 1063-1071.
- Liu, H. X., Zhang, M., and Krainer, A. R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev* 12, 1998-2012.
- Lopez, A. J. (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* 32, 279-305.
- Lou, H., Neugebauer, K. M., Gagel, R. F., and Berget, S. M. (1998). Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol Cell Biol* 18, 4977-4985.

- Luo, M. J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci U S A* 96, 14937-14942.
- Lutz, C. S., and Alwine, J. C. (1994). Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. *Genes Dev* 8, 576-586.
- Lykke-Andersen, J., Shu, M. D., and Steitz, J. A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* 293, 1836-1839.
- Lynch, K. W., and Maniatis, T. (1996). Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer. *Genes Dev* 10, 2089-2101.
- Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* 416, 499-506.
- Manley, J. L., and Tacke, R. (1996). SR proteins and splicing control. *Genes Dev* 10, 1569-1579.
- Markovtsov, V., Nikolic, J. M., Goldman, J. A., Turck, C. W., Chou, M. Y., and Black, D. L. (2000). Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol Cell Biol* 20, 7463-7479.
- Maroney, P. A., Romfo, C. M., and Nilsen, T. W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Mol Cell* 6, 317-328.
- Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M. Q., Gardiner, E. M., and Krainer, A. R. (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *Embo J* 18, 4560-4570.

- Mayeda, A., and Krainer, A. R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68, 365-375.
- Mayeda, A., and Krainer, A. R. (1999). Mammalian in vitro splicing assays. *Methods Mol Biol* 118, 315-321.
- Mayes, A. E., Verdone, L., Legrain, P., and Beggs, J. D. (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *Embo J* 18, 4321-4331.
- McCullough, A. J., and Berget, S. M. (1997). G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol Cell Biol* 17, 4562-4571.
- McCullough, A. J., and Schuler, M. A. (1997). Intronic and exonic sequences modulate 5' splice site selection in plant nuclei. *Nucleic Acids Res* 25, 1071-1077.
- McGarvey, T., Rosonina, E., McCracken, S., Li, Q., Arnaout, R., Mientjes, E., Nickerson, J. A., Awrey, D., Greenblatt, J., Grosveld, G., and Blencowe, B. J. (2000). The acute myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with exon-product complexes. *J Cell Biol* 150, 309-320.
- Michael, W. M., Choi, M., and Dreyfuss, G. (1995). A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* 83, 415-422.
- Misteli, T., Caceres, J. F., and Spector, D. L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387, 523-527.
- Misteli, T., and Spector, D. L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell* 3, 697-705.
- Modrek, B., and Lee, C. J. (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet* 34, 177-180.

- Moore, M. J., Query, C. C., and Sharp, P. A. (1993). Splicing of Precursors to mRNA by the Spliceosome. In *RNA World*, R. F. Gesteland., and J. F. Atkins., eds. (New York, CSH Press), pp. 303_357.
- Mullen, M. P., Smith, C. W., Patton, J. G., and Nadal-Ginard, B. (1991). Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev* 5, 642-655.
- Muro, A. F., Caputi, M., Pariyarath, R., Pagani, F., Buratti, E., and Baralle, F. E. (1999). Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol Cell Biol* 19, 2657-2671.
- Muro, A. F., Iaconcig, A., and Baralle, F. E. (1998). Regulation of the fibronectin EDA exon alternative splicing. Cooperative role of the exonic enhancer element and the 5' splicing site. *FEBS Lett* 437, 137-141.
- Nagel, R. J., Lancaster, A. M., and Zahler, A. M. (1998). Specific binding of an exonic splicing enhancer by the pre-mRNA splicing factor SRp55. *Rna* 4, 11-23.
- Nasim, F. H., Spears, P. A., Hoffmann, H. M., Kuo, H. C., and Grabowski, P. J. (1990). A Sequential splicing mechanism promotes selection of an optimal exon by repositioning a downstream 5' splice site in preprotachykinin pre- mRNA. *Genes Dev* 4, 1172-1184.
- Nesic, D., Cheng, J., and Maquat, L. E. (1993). Sequences within the last intron function in RNA 3'-end formation in cultured cells. *Mol Cell Biol* 13, 3359-3369.
- Niksic, M., Romano, M., Buratti, E., Pagani, F., and Baralle, F. E. (1999). Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. *Hum Mol Genet* 8, 2339-2349.
- Niwa, M., and Berget, S. M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev* 5, 2086-2095.

- Niwa, M., Rose, S. D., and Berget, S. M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev* 4, 1552-1559.
- Norton, P. A. (1994). Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding. *Nucleic Acids Res* 22, 3854-3860.
- Ou, S. H., Wu, F., Harrich, D., Garcia-Martinez, L. F., and Gaynor, R. B. (1995). Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J Virol* 69, 3584-3596.
- Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F., and Sharp, P. A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* 225, 898-903.
- Pagani, F., Buratti, E., Stuani, C., and Baralle, F. E. (2003). Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J Biol Chem* 278, 26580-26588.
- Pagani, F., Buratti, E., Stuani, C., Romano, M., Zuccato, E., Niksic, M., Giglio, L., Faraguna, D., and Baralle, F. E. (2000). Splicing factors induce CFTR exon 9 skipping through a non-evolutionary conserved intronic element. *J Biol Chem* 275, 21041-21047.
- Proudfoot, N. J., Furger, A., and Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell* 108, 501-512.
- Puig, O., Gottschalk, A., Fabrizio, P., and Seraphin, B. (1999). Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes Dev* 13, 569-580.
- Raker, V. A., Plessel, G., and Luhrmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *Embo J* 15, 2256-2269.

- Ramchatesingh, J., Zahler, A. M., Neugebauer, K. M., Roth, M. B., and Cooper, T. A. (1995). A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol Cell Biol* 15, 4898-4907.
- Reed (1997). Spliceosome assembly. In *Eukaryotic mRNA Processing*, O. I. P. Krainer AR, ed., pp. 103-129.
- Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. *Genes Dev* 3, 2113-2123.
- Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev* 6, 215-220.
- Reed, R., and Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell* 41, 95-105.
- Reed, R., and Maniatis, T. (1988). The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev* 2, 1268-1276.
- Robberson, B. L., Cote, G. J., and Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 10, 84-94.
- Rodrigues, J. P., Rode, M., Gatfield, D., Blencowe, B. J., Carmo-Fonseca, M., and Izaurralde, E. (2001). REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc Natl Acad Sci U S A* 98, 1030-1035.
- Romano, M., Marcucci, R., Buratti, E., Ayala, Y. M., Sebastio, G., and Baralle, F. E. (2002). Regulation of 3' splice site selection in the 844ins68 polymorphism of the cystathionine b-synthase gene. *J Biol Chem* 277, 12.
- Roscigno, R. F., Weiner, M., and Garcia-Blanco, M. A. (1993). A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *J Biol Chem* 268, 11222-11229.

- Rowen, L., Young, J., Birditt, B., Kaur, A., Madan, A., Philipps, D. L., Qin, S., Minx, P., Wilson, R. K., Hood, L., and Graveley, B. R. (2002). Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics* 79, 587-597.
- Ruskin, B., Greene, J. M., and Green, M. R. (1985). Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell* 41, 833-844.
- Ryner, L. C., and Baker, B. S. (1991). Regulation of doublesex pre-mRNA processing occurs by 3'-splice site activation. *Genes Dev* 5, 2071-2085.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning - A Laboratory Manual*, 2nd edn, Cold Spring Harbor Laboratory Press).
- Schaal, T. D., and Maniatis, T. (1999a). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol Cell Biol* 19, 261-273.
- Schaal, T. D., and Maniatis, T. (1999b). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol* 19, 1705-1719.
- Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E., and Zipursky, S. L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671-684.
- Scott, J., Knott, T. J., Priestley, L. M., Robertson, M. E., Mann, D. V., Kostner, G., Miller, G. J., and Miller, N. E. (1985). High-density lipoprotein composition is altered by a common DNA polymorphism adjacent to apoprotein AII gene in man. *Lancet* 1, 771-773.
- Shapiro, M. B., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 15, 7155-7174.

- Shelley, C. S., and Baralle, F. E. (1987). Deletion analysis of a unique 3' splice site indicates that alternating guanine and thymine residues represent an efficient splicing signal. *Nucleic Acids Res* 15, 3787-3799.
- Shelley, C. S., Sharpe, C. R., Baralle, F. E., and Shoulders, C. C. (1985). Comparison of the human apolipoprotein genes. Apo AII presents a unique functional intron-exon junction. *J Mol Biol* 186, 43-51.
- Shi, H., Hoffman, B. E., and Lis, J. T. (1997). A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. *Mol Cell Biol* 17, 2649-2657.
- Shin, D., Park, S., and Park, C. (2003). A splice variant acquiring an extra transcript leader region decreases the translation of glutamine synthetase gene. *Biochem J* 374, 175-184.
- Singh, R., Valcarcel, J., and Green, M. R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* 268, 1173-1176.
- Sironi, M., Cagliani, R., Pozzoli, U., Bardoni, A., Comi, G. P., Giorda, R., and Bresolin, N. (2002). The dystrophin gene is alternatively spliced throughout its coding sequence. *FEBS Lett* 517, 163-166.
- Sleeman, J. E., and Lamond, A. I. (1999). Nuclear organization of pre-mRNA splicing factors. *Curr Opin Cell Biol* 11, 372-377.
- Smith, C. W., Porro, E. B., Patton, J. G., and Nadal-Ginard, B. (1989). Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature* 342, 243-247.
- Smith, C. W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* 25, 381-388.
- Sontheimer, E. J., Sun, S., and Piccirilli, J. A. (1997). Metal ion catalysis during splicing of premessenger RNA. *Nature* 388, 801-805.

- Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315-326.
- Steitz, T. A., and Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* 90, 6498-6502.
- Sterner, D. A., Carlo, T., and Berget, S. M. (1996). Architectural limits on split genes. *Proc Natl Acad Sci U S A* 93, 15081-15085.
- Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal pentanRNP. *Mol Cell* 9, 31-44.
- Sun, Q., Mayeda, A., Hampson, R. K., Krainer, A. R., and Rottman, F. M. (1993). General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes Dev* 7, 2598-2608.
- Tacke, R., Chen, Y., and Manley, J. L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc Natl Acad Sci U S A* 94, 1148-1153.
- Tacke, R., and Manley, J. L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *Embo J* 14, 3540-3551.
- Talerico, M., and Berget, S. M. (1990). Effect of 5' splice site mutations on splicing of the preceding intron. *Mol Cell Biol* 10, 6299-6305.
- Talerico, M., and Berget, S. M. (1994). Intron definition in splicing of small *Drosophila* introns. *Mol Cell Biol* 14, 3434-3445.
- Tanaka, K., Watakabe, A., and Shimura, Y. (1994). Polypurine sequences within a downstream exon function as a splicing enhancer. *Mol Cell Biol* 14, 1347-1354.
- Tautz, D., and Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res* 12, 4127-4138.

- Teigelkamp, S., Achsel, T., Mundt, C., Gothel, S. F., Cronshagen, U., Lane, W. S., Marahiel, M., and Luhrmann, R. (1998). The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *Rna* 4, 127-141.
- Tian, H., and Kole, R. (1995). Selection of novel exon recognition elements from a pool of random sequences. *Mol Cell Biol* 15, 6291-6298.
- Tian, M., and Maniatis, T. (1992). Positive control of pre-mRNA splicing in vitro. *Science* 256, 237-240.
- Tian, M., and Maniatis, T. (1993). A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* 74, 105-114.
- Tian, M., and Maniatis, T. (1994). A splicing enhancer exhibits both constitutive and regulated activities. *Genes Dev* 8, 1703-1712.
- Tran, Q., Coleman, T. P., and Roesser, J. R. (2003). Human transformer 2beta and SRp55 interact with a calcitonin-specific splice enhancer. *Biochim Biophys Acta* 1625, 141-152.
- Tran, Q., and Roesser, J. R. (2003). SRp55 is a regulator of calcitonin/CGRP alternative RNA splicing. *Biochemistry* 42, 951-957.
- Tsukahara, T., Casciato, C., and Helfman, D. M. (1994). Alternative splicing of beta-tropomyosin pre-mRNA: multiple cis- elements can contribute to the use of the 5'- and 3'-splice sites of the nonmuscle/smooth muscle exon 6. *Nucleic Acids Res* 22, 2318-2325.
- Umen, J. G., and Guthrie, C. (1995). The second catalytic step of pre-mRNA splicing. *Rna* 1, 869-885.
- Valcarcel, J., and Green, M. R. (1996). The SR protein family: pleiotropic functions in pre-mRNA splicing. *Trends Biochem Sci* 21, 296-301.

- Visa, N., Alzhanova-Ericsson, A. T., Sun, X., Kiseleva, E., Bjorkroth, B., Wurtz, T., and Daneholt, B. (1996). A pre-mRNA-binding protein accompanies the RNA from the gene through the nuclear pores and into polysomes. *Cell* 84, 253-264.
- Wagner, E. J., Carstens, R. P., and Garcia-Blanco, M. A. (1999). A novel isoform ratio switch of the polypyrimidine tract binding protein. *Electrophoresis* 20, 1082-1086.
- Webb, K. E., Martin, J. F., Cotton, J., Erusalimsky, J. D., and Humphries, S. E. (2003). The 4830C>A polymorphism within intron 5 affects the pattern of alternative splicing occurring within exon 6 of the thrombopoietin gene. *Exp Hematol* 31, 488-494.
- Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44, 388-396.
- Will, C. L. (1997). snRNP structure and function. In *Eukaryotic mRNA processing*, e. O.I.P. Krainer AR, ed. (Oxford), pp. 130-173.
- Wittig, B., Wolfl, S., Dorbic, T., Vahrson, W., and Rich, A. (1992). Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *Embo J* 11, 4653-4663.
- Wittop Koning, T. H., and Schumperli, D. (1994). RNAs and ribonucleoproteins in recognition and catalysis. *Eur J Biochem* 219, 25-42.
- Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75, 1061-1070.
- Xiao, S. H., and Manley, J. L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* 11, 334-344.
- Xu, R., Teng, J., and Cooper, T. A. (1993). The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol Cell Biol* 13, 3660-3674.

- Yean, S. L., Wuenschell, G., Termini, J., and Lin, R. J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* 408, 881-884.
- Yitzhaki, S., Miriami, E., Sperling, R., and Sperling, J. (1996). Phosphorylated Ser/Arg-rich proteins: limiting factors in the assembly of 200S large nuclear ribonucleoprotein particles. *Proc Natl Acad Sci U S A* 93, 8830-8835.
- Yu, Y. T., Maroney, P. A., Darzynkiwicz, E., and Nilsen, T. W. (1995). U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone. *Rna* 1, 46-54.
- Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R. V., Gentile, C., Gebara, M., and Corden, J. L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A* 93, 6975-6980.
- Zamore, P. D., and Green, M. R. (1991). Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *Embo J* 10, 207-214.
- Zamore, P. D., Patton, J. G., and Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355, 609-614.
- Zeng, C., Kim, E., Warren, S. L., and Berget, S. M. (1997). Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *Embo J* 16, 1401-1412.
- Zhang, D., and Rosbash, M. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev* 13, 581-592.
- Zheng, Z. M., He, P. J., and Baker, C. C. (1997). Structural, functional, and protein binding analyses of bovine papillomavirus type 1 exonic splicing enhancers. *J Virol* 71, 9096-9107.

- Zheng, Z. M., Quintero, J., Reid, E. S., Gocke, C., and Baker, C. C. (2000). Optimization of a weak 3' splice site counteracts the function of a bovine papillomavirus type 1 exonic splicing suppressor in vitro and in vivo. *J Virol* 74, 5902-5910.
- Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* 407, 401-405.
- Zhu J, M. A., Krainer A (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 8, 1351-1361.
- Zhuang, Y., and Weiner, A. M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev* 3, 1545-1552.
- Zuo, P., and Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev* 10, 1356-1368.
- Zuo, P., and Manley, J. L. (1994). The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. *Proc Natl Acad Sci U S A* 91, 3363-3367.